NATIONAL QUALITY FORUM

# MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

**To navigate the links in the worksheet: Click to go to the link. ALT + LEFT ARROW to return**

**Purple** text represents the responses from measure developers.

**Red** text denotes developer information that has changed since the last measure evaluation review.

## Brief Measure Information

**NQF #:** 2633

**Corresponding Measures:**

**De.2. Measure Title:** Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Change in Self-Care Score for Medical Rehabilitation Patients

**Co.1.1. Measure Steward:** Centers for Medicare & Medicaid Services

**De.3. Brief Description of Measure:** This measure estimates the risk-adjusted mean change in self-care score between admission and discharge for Inpatient Rehabilitation Facility (IRF) Medicare Part A and Medicare Advantage patients.

**1b.1. Developer Rationale:** During an Inpatient Rehabilitation Facility (IRF) stay, the goals of treatment include fostering the patient's ability to manage his or her daily activities so that the patient can complete self-care and mobility activities as independently as possible and, if feasible, return to a safe, active and productive life in a community-based setting. Given that the primary goal of rehabilitation is function improvement, IRF clinicians have traditionally assessed and documented patients' functional status at admission and discharge to calculate change in function scores. The change in function scores represent the effectiveness of the rehabilitation care provided to patients in the rehabilitation unit or hospital.

The self-care quality measure uses standardized data elements for the collection of functional status data, which can improve communication when patients are transferred between providers. Most IRF patients receive care in an acute care hospital prior to the IRF stay, and many IRF patients receive care from another provider after the IRF stay. Use of standardized clinical data to describe a patient´s status across providers can facilitate communication across providers.

In describing the importance of functional status, the National Committee on Vital and Health Statistics Subcommittee on Health (2001) noted, "Information on functional status is becoming increasing essential for fostering healthy people and a health population. Achieving optimal health and well-being for Americans requires an understanding across the life space of the effects of people's health conditions on their ability to do basic activities and participate in life situations, in other words, their functional status."

This quality measure will inform IRF providers about opportunities to improve care in the area of function and strengthen incentives for quality improvement related to patient function.

Citation:

National Committee on Vital and Health Statistics Subcommittee on Health. Classifying and Reporting Functional Status. 2001. Retrieved from http://www.ncvhs.hhs.gov/010617rp.pdf

**S.4. Numerator Statement:** The measure does not have a simple form for the numerator and denominator. This measure estimates the risk-adjusted change in self-care score between admission and discharge among Inpatient Rehabilitation Facility (IRF) Medicare Part A and Medicare Advantage patients age 21 or older. The change in self-care score is calculated as the difference between the discharge self-care score and the admission self-care score.

**S.6. Denominator Statement:** The denominator is the number of Inpatient Rehabilitation Facility Medicare patient stays, except those that meet the exclusion criteria.

**S.8. Denominator Exclusions:** This quality measure has six patient-level exclusion criteria:

1) Patients with incomplete stays.

Rationale: When a patient has an incomplete stay, for example, the patients leave urgently due to a medical emergency, it can be challenging to gather accurate discharge functional status data. Patients with incomplete stays include patients who are unexpectedly discharged to an acute care setting (Short-stay Acute Hospital, Critical Access Hospital, Inpatient Psychiatric Facility, or Long-term Care Hospital); patients who die or leave an Inpatient Rehabilitation Facility (IRF) against medical advice; and patients with a length of stay less than 3 days.

2) Patients who are independent with all self-care activities at the time of admission.

Rationale: Patients who are independent with all the self-care items at the time of admission are assigned the highest score on all the self-care items, and thus, would not be able to show functional improvement on this same set of items at discharge.

3) Patients with the following medical conditions on admission: coma; persistent vegetative state; complete quadriplegia; locked-in syndrome; or severe anoxic brain damage, cerebral edema or compression of the brain.

Rationale: These patients are excluded because they may have limited or less predictable self-care improvement with the selected self-care items.

4) Patients younger than age 21.

Rationale: There is only limited evidence published about functional outcomes for individuals with Medicare who are younger than 21.

5) Patients discharged to Hospice.

Rationale: Patient goals may change during the IRF stay, and functional improvement may no longer be a goal for a patient discharged to hospice.

6) Patients who are not Medicare Part A and Medicare Advantage beneficiaries.

Rationale: IRF-PAI data for patients not covered by the Medicare program are not submitted to the Centers for Medicare and Medicaid Services.

Facility-level quality measure exclusion: For IRFs with fewer than 20 patient stays, data for this quality measure are not publicly reported.

**De.1. Measure Type:** Outcome

**S.17. Data Source:** Instrument-Based Data

**S.20. Level of Analysis:** Facility

**IF Endorsement Maintenance – Original Endorsement Date:** Nov 04, 2015 **Most Recent Endorsement Date:** Nov 04, 2015

**IF this measure is included in a composite, NQF Composite#/title:**

**IF this measure is paired/grouped, NQF#/title:**

**De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** Not applicable. This measure is not paired or grouped with another measure.

## Preliminary Analysis: Maintenance of Endorsement

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

## Criteria 1: Importance to Measure and Report

### 1a. Evidence

**Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.**

**1a. Evidence.** The evidence requirements for a health outcome measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance, assuming the data are from a robust number of providers and results are not subject to systematic bias. For measures derived from patient report, evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

**Evidence Summary**

- Three studies show better functional outcomes for patients for IRF patients compared to other post-acute care settings, likely because IRFs provide the most intensive therapy services of the PACs.
- Similar patients tend to be treated in SNFs, and research has demonstrated SNF patients receiving therapy tended to have improved functional outcomes, with greater improvement for more hours of therapy.

**Changes to evidence from last review**

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☒ **The developer provided updated evidence for this measure:**

**Updates:**

- Developer offered a logic model depicting the relationship between structures, processes and patient outcomes.
- Developer conducted scoping review for literature examining the relationship between therapy interventions and improved patient function, published since January 1, 2013 (after last submission).
- Most IRF research on functional outcomes focused on motor function, encompassing self-care and mobility, and sometimes bladder function. Developer states several observational studies reported positive associations between amount of therapy provided and motor function. Studies also found older patients tended to receive fewer hours of treatment and overall regained less function, and that frailer patients were less likely to regain baseline functional ability.
- Developer states that rehab interventions tend to be multidisciplinary and tailored to individual patients which makes it challenging to examine specific interventions. However, two IRF studies found additional interventions added to "usual" therapy generally improved functional or motor outcomes.
- Three studies show better functional outcomes for patients for IRF patients compared to other post-acute care settings, likely because IRFs provide the most intensive therapy services of the PACs.

- Similar patients tend to be treated in SNFs, and research has demonstrated SNF patients receiving therapy tended to have improved functional outcomes, with greater improvement for more hours of therapy.

***Question for the Committee:***

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

**Guidance from the Evidence Algorithm**

Measure assesses outcome (box 1) YES -> relationship between outcome and at least one healthcare action (box 2) YES -> PASS

**Preliminary rating for evidence: ☒ Pass ☐ No Pass**

## 1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

**Maintenance measures – increased emphasis on gap and variation**

**1b. Performance Gap.** The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- Quality measure score distributions over 12-months were similar between fiscal year 2017 (mean: 11.4; standard deviation: 1.7) and between calendar year 2017 (mean: 11.5; standard deviation: 1.7).
- Quality measure scores by decile show variations in quality measure scores across IRFs. The interquartile range for the two periods was 2.2 self-care units.
- Across five quarters (Q4, 2016 – Q4, 2017), mean scores increased marginally from 11.3 to 11.5 and quality measure score distributions showed variation in IRF outcomes.

**Disparities**

- The mean unadjusted (observed) change in self-care score varied slightly by dual eligibility status, race, and living alone status. Dual eligible patients with full Medicaid benefits had on average 11.0 units of change in self-care while patients who were dual eligible without full Medicaid benefits or who were non-dual eligible had more change in self-care (12.0 and 11.6 units, respectively). For race, the highest mean change in self-care was found among patients who were white (11.6 units of change), multiracial (11.5 units of change), or Native American or Alaskan Native (11.4 units of change) whereas the lowest was among patients who were Asian (10.4 units of change). Patients who were living alone prior to their hospitalization had on average 12.0 units of change in self-care whereas those not living alone had 11.3 units of change in self-care. The mean unadjusted (observed) change in self-care scores were similar across Hispanic ethnicity, urbanicity, and SES.
- The testing of social risk factors and their relationships to patients' change in self-care scores indicate that some factors (full dual eligibility, Black, Asian or Native Hawaiian race) were tied to slightly lower self-care change scores while others (lower SES, living alone, Hispanic ethnicity) were tied to higher self-care change scores.
- Multiple studies have shown that IRF patients' functional outcomes differ by geographic region, facility characteristics, IRF length of stay, and race/ethnicity, after adjusting for key demographic characteristics and admission clinical status.

***Questions for the Committee:***

- Does the performance gap reported by the developer represent opportunity for improvement that can be addressed by the measured entities?

**Preliminary rating for opportunity for improvement: ☐ High ☒ Moderate ☐ Low ☐ Insufficient**

**Committee Pre-evaluation Comments:**
**Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)**

**1a. Evidence**: *For all measures (structure, process, outcome, patient-reported structure/process), empirical data are required. How does the evidence relate to the specific structure, process, or outcome being measured? Does it apply directly or is it tangential? How does the structure, process, or outcome relate to desired outcomes? For maintenance measures –are you aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission? For measures derived from a patient report: Measures derived from a patient report must demonstrate that the target population values the measured outcome, process, or structure.*

- The developer provided empirical data that shows improved patient outcomes in IRFs, showing an association with intensity of therapy offered in these settings. Limitations of the evidence are highlighted by the developer and indicate most interventions are multidisciplinary and tailored to meet an individual's needs, which makes it difficult to assess specific interventions. In other instances, the evidence is related to functional outcomes such as those focused on motor function, encompassing self-care and mobility, and sometimes bladder function. Three studies show better functional outcomes for patients for IRF patients compared to other post-acute care settings, likely because IRFs provide the most intensive therapy services of the Post-Acute Care settings (PACs).
- Evidence supports measure
- For all measures (structure, process, outcome, patient-reported structure/process), empirical data are required. How does the evidence relate to the specific structure, process, or outcome being measured? Provided in the analysis supplied by CMS. Does it apply directly or is it tangential? This applies directly. How does the structure, process, or outcome relate to desired outcomes? Unclear if the "outcome" is return to independence or something else. This is not clearly defined. For maintenance measures –are you aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission? I found a large number of studies (~900) on a Pub Med Search including some from 2019 using the search strategy "fim score, inpatient rehabilitatation". An addition reference was also noted in Appendix C page 59 of the Competing Measures Memo of June 11, 2019: Citation: Fisher, Steve R., Middleton, Addie, Graham, James E., Ottenbacher, Kenneth J.. (2018). Same but different: FIM summary scores may mask variability in physical functioning profiles. Archives of Physical Medicine and Rehabilitation, 99(8), Pgs. 1479-1482, 1482.e1. Retrieved 12/6/2018, from REHABDATA database. The references in this Memo provided on page 13 are at least 10 years old. This is a disappointment from my perspective. For measures derived from a patient report: Measures derived from a patient report must demonstrate that the target population values the measured outcome, process, or structure." The measure developers did not provide any evaluation or data regarding the point of view of patients who are evaluated using the patient reported outcomes within this particular measure.

**1b. Performance Gap**: *Was current performance data on the measure provided? How does it demonstrate a gap in care (variability or overall less than optimal performance) to warrant a national performance measure? Disparities: Was data on the measure by population subgroups provided? How does it demonstrate disparities in the care?*

- Yes, the developer provided performance gap in care data. The summary highlighted data from various studies indicating variations across select racial and ethnic groups (full dual eligibility, Black, Asian or Native Hawaiian races), geographic areas, length of stay, and facility characteristics. These gaps represent opportunities for ongoing measurement and improvement.
- Variations in scores across IRFs is noted, does demonstrate need for a national performance measure. Disparities are discussed and a potential for improvement with the use of the measure is demonstrated.
- 4. 1b. Performance Gap: Was current performance data on the measure provided? Yes. How does it demonstrate a gap in care (variability or overall less than optimal performance) to warrant a national performance measure? Unclear. The most currently available data does not show significant change within deciles over a recent 2 year period suggesting that this measure may be "topped out". Disparities: Was data on the measure by population subgroups provided? Not that I could find, which is disappointing given the heterogeneity of this particular population and its various subpopulations. How does it

demonstrate disparities in the care? Not provided. This was noted in the 2015 Standing Committee voting summary, though, on page 65 of the Memo: "The Committee inquired about the lack of information on disparities in measure performance; the developer indicated the data is available; however, due to the wealth of information they have, they were unsure how much and what data to submit. They agreed to provide additional information, specifically on age, race and payer source, during the public comment period." No additional information on this topic was included in the current measure summary data for either 2286 or 2633.

## Criteria 2: Scientific Acceptability of Measure Properties

**2a. Reliability:** Specifications **and** Testing

**2b. Validity:** Testing**;** Exclusions**;** Risk-Adjustment**;** Meaningful Differences**;** Comparability**;** Missing Data

### Reliability

**2a1. Specifications** requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

**2a2. Reliability testing** demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers. For maintenance measures – less emphasis if no new testing data provided.

### Validity

**2b2. Validity testing** should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For maintenance measures – less emphasis if no new testing data provided.

**2b2-2b6. Potential threats to validity** should be assessed/addressed.

**Complex measure evaluated by Scientific Methods Panel**? ☒ **Yes** ☐ **No**

**Evaluators:** NQF Scientific Methods Panel

Methods Panel Review (Combined)

**Methods Panel Evaluation Summary**:

This measure was reviewed by the Scientific Methods Panel and discussed on the call. A summary of the measure and the Panel discussion is provided below.

**Scientific Methods Panel Votes: Measure Passes**

- Reliability: H-4, M-2, L-0, I-0
- Validity: H-2, M-3, L-1, I-0

**Standing Committee Summary**

The NQF Scientific Methods Panel reviewed this measure and elected not to discuss it further based on achieving a consensus that the measure should pass to the Standing Committee.

Reliability

- Performance score level:
    - Split-half reliability (note max change score is 35 points/client)
    - Results: r=0.90, ICC=.90, p<.0001; not appreciably different for providers with different volumes (Table 3)

- Data element level
    - Results: interrater reliability on items, adjusted, had Kappa ranges from 0.63 (toileting) to 0.58 (oral hygiene) (n=34 providers across 11 geo-areas, each with at least 20 patients).
    - Internal consistency: Cronbach's alpha= 0.94

Validity

- Table 4 shows good correspondence between self-care instrument items and 7 other instruments.
- 57% of TEP rated scientific soundness as high or moderate
- Patients with higher scores were more likely to be discharged to community (versus more medically intensive) settings. The was evident in a logical and 'dose-response' way from dependent to independent for all individual items, except slight deviations from that pattern for eating. (Table 5)
- Logistic regression: "show that, on average, a one-unit increase in discharge self-care score is associated with a 16 percent increase in the odds of being discharged to the community (OR = 1.159; *p*-value <0.001)."
- Rasch analysis shows that harder activity (hardest: footwear dressing) correlates with higher self-care scores than the easier activity (easiest: eating). Fit statistics also generally supportive, though not fully understandable from narrative alone.
- Table 9: Higher IRF scores correlated with higher achievement of certification as a stroke rehab center (Table 9). Link to quality. Mean change in self-care performance score ~5%. (Table 10 with quintiles more persuasive). This analysis could be strengthened by just testing stroke cases (25% of all cases studied by developers own account).
- Exclusion analysis: 11% (55K) incomplete stays; only 32 patients were <21 years old; 0.5% to hospice; 0.1% were fully independent per scale upon admission.
- Statistically meaningful differences: Three performance groups each >350 providers: 95% CI below, overlapping, or above mean. CIs not presented.
- Missing data from discharge or admission <0.01% across all 7 self-care data elements across the 427K observations.

Standing Committee Action Items:

- The Standing Committee can discuss reliability and/or validity or accept the Scientific Methods Panel ratings.
- The Standing Committee may also elect to discuss the SMP's identified concerns on validity:
    - Concerns about the following exclusions:
        - Patients with incomplete stays.
        - Patients discharged to Hospice.
        - Patients who are not Medicare Part A and Medicare Advantage beneficiaries.
    - Discharge score may be used to determine destination, so perhaps not good to use destination to community as a validity check?
    - Concerns that TEP approval of the instrument was not higher than 57%?
    - Correspondence between measure score and stroke facility certification was evident, but small (<5% difference in mean difference scored).
    - Confirm with developer that score level reliability and validity was risk-adjusted with their acuity model.

***Questions for the Committee regarding reliability:***

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- The Scientific Methods Panel is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?

***Questions for the Committee regarding validity:***

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- The Scientific Methods Panel is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?

**Preliminary rating for reliability:** ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

**Preliminary rating for validity:** ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

## Combined Methods Panel Scientific Acceptability Evaluation

**Measure Number:** 2633

**Measure Title: Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Change in Self-Care Score for Medical Rehabilitation Patients**

### Type of measure:

☐ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use**

☒ **Outcome** ☐ **Outcome: PRO-PM** ☐ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**

### Data Source:

☐ **Claims** ☐ **Electronic Health Data** ☐ **Electronic Health Records** ☐ **Management Data**

☒☐ **Assessment Data** ☐ **Paper Medical Records** ☐☒ **Instrument-Based Data** ☐ **Registry Data**

☐ **Enrollment Data** ☐☐ **Other MP 1:**CARE Tool used during the Post-Acute Care Payment Reform Demonstration – electronic submission. IRF-PAI) **MP 2 and MP 3:**Inpatient Rehabilitation Facility Patient Assessment Instrument (IRF-PAI)

### Level of Analysis:

☐ **Clinician: Group/Practice** ☐ **Clinician: Individual** ☒ **Facility** ☐ **Health Plan**

☐ **Population: Community, County or City** ☐ **Population: Regional and State**

☐ **Integrated Delivery System** ☐ **Other**

### Measure is:

☐☐ **New** ☐☒ **Previously endorsed (**NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.)

**MP 1:**The measure does not have a simple form for the numerator and denominator. This measure estimates the risk-adjusted change in self-care score between admission and discharge among Inpatient Rehabilitation Facility (IRF) Medicare Part A and Medicare Advantage patients age 21 or older. The change in self-care score is calculated as the difference between the discharge self-care score and the admission self-care score.

The denominator is the number of Inpatient Rehabilitation Facility Medicare patient stays, except those that meet one or more of the six the exclusion criteria.

### RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?** ☒ **Yes** ☐ **No**

   **Submission document:** "MIF_xxxx" document, items S.1-S.22

   ***NOTE****: NQF staff will conduct a separate, more technical, check of eCQM specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*

2. **Briefly summarize any concerns about the measure specifications.**

   **MP 2:**No Concerns

   **MP 4:**None

### RELIABILITY: TESTING

**Submission document:** "MIF_xxxx" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

3. **Reliability testing level** ☐☒ **Measure score** ☒ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure** ☒ **Yes** ☐ **No**
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical <u>VALIDITY</u> testing** of <u>patient-level data</u> conducted?

   ☐ **Yes** ☐ **No** **MP 4:**NA

6. **Assess the method(s) used for reliability testing**

   **Submission document:** Testing attachment, section 2a2.2

   **MP 2:**Split-half reliability was used to examine the reliability of the computed performance measure scores. The Pearson Product-Moment Correlation (r), Spearman Rank Correlation (ρ), and Intraclass Correlation Coefficient (ICC) were used to examine the performance measure reliability. Intraclass correlations were also calculated by facility volume quartile to examine whether there were differences in performance measure reliability by IRF size. Internal consistency was assessed using the Cronbach's alpha coefficient. Inter-Rater Reliability based on previous studies

   **MP 1:**

   a. Computed Performance Measure Score Reliability – Split-half Reliability (unit of analysis is providers): Split-half reliability was used to examine the reliability of the computed performance measure scores. The computed performance measure scores are the risk-adjusted change in self-care scores. Only facilities with 20 or more stays were included. Developer conducted split-half reliability by randomly splitting each provider's patient stays into two groups and calculating correlations between the computed performance measure scores of the randomly divided groups. The Pearson Product-Moment Correlation (*r*), Spearman Rank Correlation (*ρ)*, and Intraclass Correlation Coefficient (ICC) were used to measure internal reliability. Intraclass correlations were also calculated by facility volume quartile to examine whether there were differences in performance measure reliability by IRF size
   b. Self-Care Scale/Instrument Analysis - Internal Consistency (unit of analysis is patient assessments): In addition to the provider-level reliability testing of the computed performance measure scores described above, developer examined the internal consistency of the self-care scale/instrument scores for each patient-stay.
   c. Internal consistency was assessed using the Cronbach's alpha
   d. Critical Data Elements Testing using CARE Tool Data (2014) – Inter-Rater Reliability, Video (Standardized Patient) Reliability and Validity Testing (unit of analysis is patients): Reliability and validity testing included the self-care and mobility data elements, as well as data elements that are used as risk adjustors for this performance measure

   **Methods were reasonable**

   **MP 4:**Appropriate.

**MP 3:** The developer conducted reliability testing for both data element and measure score. For data element reliability, the developer reported internal consistency, inter-rater reliability (this is relevant for the mobility score is assessed by clinicains using the instrument). In addition, the developer reported the results

from the video reliability study. For measure score reliability, the developer conducted split-half reliability testing. Both methods were appropriate.

**MP 5:**The developer used Cronbach's alpha to evaluate internal consistency of the self-care items and used an ICC with a split half approach to evaluate facility level reliability. The developer used risk-adjusted scores in the ICC analysis. The developer also conducted inter-rater reliability of the data elements. All of the approaches were appropriate.

**MP 6:**For data element reliability, a Cronbach's alpha measure of internal (consistent was used – this is a standard method for multi-item scales, although a little unusual when the items are intentionally designed to be scaled in their difficulty. It's not clear how high the Cronbach's alpha results.SHOULD be if it expected that a patient at a given level of self-care will be able to do some things independently, other things with assistance, and perhaps not be able to do some things at all. For measure score reliability, a split-half approach was used in which the cases for each facility were divided in half, with scores derived from each half, and then the results compared for the two half-samples, using three different statistical tests.

7. **Assess the results of reliability testing**

**Submission document:** Testing attachment, section 2a2.3

**MP 2:**Previous IRR demonstrated acceptable Kappa scores. ): Split-half analysis results indicated positive moderate-to-strong correlations. ICCs for the volume quartiles showed moderate to strong scores.

**MP 3:** Cronbach's alphas was very good, inter-rater reliability measured by weighted kappa was also substantial. In general, results for data element reliabity testing were very good.

Measure score reliability measured by ICC was also quite high.

**MP 4:**Adequate test sample. Moderate to high confidence regarding reliability of measure results.

**MP 1:**Weighted kappa values for the self-care items range between 0.798 for eating to 0.869 for upper-body dressing. Unweighted kappas ranged from 0.598 for oral hygiene to 0.634 for upper-body dressing. Unweighted overall kappas ranged from 0.636 (toileting) to 0.598 (oral hygiene). Kappa statistics indicated sufficient agreement of data element codes among raters.
For the video reliability study, clinicians assessed "standardized" patients presented through a videotape of a patient assessment. This ensured that the same information was presented to each clinician and allowed examination of scoring among different clinicians examining the "same" patient. The video reliability study indicated substantial agreement with the mode and clinical team among all items, typically upwards of 70%. The notable exception to this trend exists among the clinicians in the "Other" category (mostly LPNs); they consistently had the lowest levels of agreement among self-care items, ranging from 50 to 72%. For the toileting and dressing items, the agreement with the clinical team was lower than with the mode CWeighted kappa values for the self-care items range between 0.798 for eating to 0.869 for upper-body dressing. Unweighted kappas ranged from 0.598 for oral hygiene to 0.634 for upper-body dressing. Unweighted overall kappas ranged from 0.636 (toileting) to 0.598 (oral hygiene). Kappa statistics indicated sufficient agreement of data element codes among raters.

For the video reliability study, clinicians assessed "standardized" patients presented through a videotape of a patient assessment. This ensured that the same information was presented to each clinician and allowed examination of scoring among different clinicians examining the "same" patient. The video reliability study indicated substantial agreement with the mode and clinical team among all items, typically upwards of 70%. The notable exception to this trend exists among the clinicians in the "Other" category (mostly LPNs); they consistently had the lowest levels of agreement among self-care items, ranging from 50 to 72%. For the toileting and dressing items, the agreement with the clinical team was lower than with the mode

MP 5: Cronbach's alpha was high for the self-care data elements (alpha = 0.94). The facility-level reliability estimates were consistently strong regardless of discharge volume at the facility level. For the full sample, the reliability coefficient was 0.91, which is very good,. Inter-rater reliabilityof the self-care data elements was good.

**MP 6:**The results of testing at both data element and measure relability showed acceptable reliability, using standard and generally-accepted methods.

8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? NOTE: If multiple methods used, at least one must be appropriate.

   **Submission document:** Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

   ☐ **Not applicable** (score-level testing was not performed)

9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

   **Submission document:** Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

   ☐ **Not applicable** (data element testing was not performed)

10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and all testing results):

    ☒☐ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

    ☐☒ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has not been conducted)

    ☐ **Low** (NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

    ☐ **Insufficient** (NOTE: Should rate INSUFFICIENT if you believe you do not have the information you need to make a rating decision)

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**

    **MP 2:**All analysis demonstrated reliability. No concerns noted

    **MP 4:**Testing results.

**MP 3:**Comprehensive reliability testing was conducted, covering both data element and measure score. The results were very good.

**MP 5:**Strong internal consistency of items and high ICC correlations at the facility level using the split half approach warrants a rating of high reliability.

**MP 6:**As noted above, the results of reliability tests were generally positive, and the measure score reliability depends (as it normally does) on having an adequate sample size.

### VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

12. **Please describe any concerns you have with measure exclusions.**

    **Submission document:** Testing attachment, section 2b2.

    **MP 2:**No concerns

**MP 1:**Six exclusions plus facility-level exclusion when fewer than 20 patients. I have concern abou the following three:

- Patients with incomplete stays.
- Patients discharged to Hospice.
- Patients who are not Medicare Part A and Medicare Advantage beneficiaries.

**MP 4:**None.

**MP 3:**No concern.

**MP 6:**None – exclusions seemed appropriate.

**MP 5:** None; exclusions appear appropriate.

13. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**

**Submission document:** Testing attachment, section 2b4.

**MP 2:**No concerns

**MP 4:**None

**MP 3:**No concern.

**MP 5:** None.

**MP 6:**Although the developers were able to show that a number of facilities had performance that was significantly above or below a national average, it is not clear whether those differences are clinically meaningful to patients or family members.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**

**Submission document:** Testing attachment, section 2b5.
**MP 2:**No concerns
**MP 4:** NA
**MP 3:**No concern.
**MP 5:** N/A

15. **Please describe any concerns you have regarding missing data.**

**Submission document:** Testing attachment, section 2b6.

**MP 2:**As noted in similar measures, concerns of scoring missing data as 01 but submitors provided analysis showing minimal missing data elements

**MP 1:No concern….missing data less than 1%**

**MP 4:**None

**MP 3:**No concern, in general very minimal missing data for all self-care data elements.

**MP 5:** None.

16. **Risk Adjustment**

16a. **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

16b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☐ Yes ☒☐ No ☒☐ Not applicable

16c. **Social risk adjustment:**

16c.1 Are social risk factors included in risk model? ☐☒ Yes ☒☐ No ☐ Not applicable

16c.2 Conceptual rationale for social risk factors included? ☐☒ Yes ☒☐ No NA

16c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☐☒ Yes ☒☐ No

16d.**Risk adjustment summary:**

16d.1 All of the risk-adjustment variables present at the start of care? ☒ Yes ☐ No

16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☒☐ Yes ☐ No NA

16d.3 Is the risk adjustment approach appropriately developed and assessed? ☐☒ Yes ☐ No

16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration)

☒ Yes ☐ No

16d.5.Appropriate risk-adjustment strategy included in the measure? ☐☒ Yes ☐ No

16e**. Assess the risk-adjustment approach**

**MP 2:**Agree with submitors that social risk facors did not impact the results so no adjustment for risk factors needed

**MP 4:**Appropriate

**MP 1:**Social factors are not included in the quality measure calculation, but the developers analyzed it anyway, and results seem reasonable

**MP 3:**Overall, the risk adjustment approach was acceptable.

**MP 5:** The risk-adjustment approach uses a total of 63 covariates which had an r-squared value of 23%. While the rsq value is excellent, even though there are quite a number of observations (n=428,192), the developers did not include any statistics to determine potential model overfit (i.e. predicted r-square).

**MP 6:**The approach was generally thoughtful and acceptable, but the developers found that some racial or ethnic groups had significantly different outcomes, and those factors could have been included in the adjustment modelts, but the developers fell back on the standard CMS "more research is necessary to understand this phenomenon" rationale for not including these social factors. Unfortunately, a decision to either include or exclude the factors has consequences for the affected providers, for patients and families, and for other stakeholders, so a decision to not include social factors like race or ethnicity because of "insufficient research" is itself a decision with potential adverse consequences for facilities serving minority patients, whose performance will appear to be worse than it actually may be. The data on the small effects of risk adjustment isn't entirely convincing, as one would expect the mean and median scores in a distribution to not change with adjustment; the key issue is how many individual facilities would move up or down by some defined amount in the distribution with adjustment. It may be true that no facilities would have moved much in the distribution with a more-inclusive adjustment model, but that information doesn't seem to have been provided.

**<u>For cost/resource use measures ONLY</u>:**

17. **Are the specifications in alignment with the stated measure intent?**

☒ **Yes** ☐ **Somewhat** ☐ **No (If "Somewhat" or "No", please explain)**

18. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

### VALIDITY: TESTING

19. **Validity testing level:** ☐☒ **Measure score** ☐☒ **Data element** ☒ ☒☐ **Both**

20. **Method of establishing validity of the measure score:**

☐☒ **Face validity**

☒ **Empirical validity testing of the measure score**

☐ **N/A (score-level testing not conducted)**

21. **Assess the method(s) for establishing validity**

**Submission document: Testing attachment, section 2b2.2**

**MP 2:**construct validity of the self-care data by examining the relation between discharge functional scores and the discharge destination. Face validity with technical expert panel. Construct validity of the self-care data by examining the relation between discharge self-care scores and being discharged to the community, after excluding incomplete stays. Scale/Instrument Construct Validity by a logistic regression

model to examine the association between observed discharge self-care scores and the odds of a community discharge. Rasch analysis uses item data to determine how well items in a scale/instrument function together to measure a construct. Content validity by comparing other self-care measurement instruments. Performance score of Joint Commission Stroke Rehab Certification compared to non-certified facilities indocated better performance score with TJC certification

**MP 1:**

a. **Scale/Instrument Content Validity - Similarity of Data Elements Across Other Self-Care Assessment Instruments:** compares favorably

b. **Face Validity -** in-person Technical Expert Panel (TEP) to seek expert input on the Inpatient Rehabilitation Facilities Quality Reporting Program (IRF QRP) quality measures, including the functional status performance measures. Prior to the TEP meeting, TEP members provided feedback on the importance, scientific soundness and usability of each of the performance measures

c. **Data Element Construct Validity – Observed Discharge Self-Care Scores and Discharge Destination (unit of analysis is patient stays):** tested the validity of the self-care data by examining the discharge function scores and whether patients were discharged to a community destination. construct validity of the self-care data suggested to come from examining the relation between discharge self-care scores and being discharged to the community, after excluding incomplete stays.

d. **Scale/Instrument Construct Validity** – **Observed Discharge Self-Care Scores and Discharge Destination (unit of analysis is patient stays)**: Examined the discharge self-care scale scores and whether patients were discharged to a community destination. Logistic regression model examined the association between discharge self-care scores and the odds of a community discharge.

e. **Scale/Instrument Construct Validity - Data Element (Item) Difficulty Ordering Using Rasch Analysis (unit of analysis is patient assessment data):** Rasch analysis uses item data to determine how well items in a scale/instrument function together to measure a construct.

f. **Data Element (Item) and Scale/Instrument Validity - Fit Assessment Analysis (unit of analysis is patient assessment data):** Rasch analysis fit statistics that reflect whether unmodeled (unexpected) responses are being coded for items within the scale/instrument. This shows up as item misfit.

g. **Data Element (Item) and Scale/Instrument Validity - Response Option Assessment Using Rasch Analysis (unit of analysis is patient assessment data):** Rasch analysis output reports the number and percent of patients by score level (06 - Independent to 01 - Dependent) for each item and the average self-care ability (i.e., scale-level ability) of those patients. This allows examination as to whether or not the 6-point rating scale works as intended.

**MP 4:**Multiple methods used to establish validity. Sufficient validity demonstrated.

**MP 3:**Face validity was assessed by polling TEP member before TEP meeting.

Extensive data element validity tests were conducted, including content validity, data elements construct validity, scale construct validity, and others. For both data elements and scale construct validity testing, the developer correlated discharge destination with both the data elements and scale and considered positive relationship as evidence of validity. It can be argued that discharge destination is not an ideal evaluation criterion as discharge destination may be partly determined by the self-care score assessed.

For measure score validity testing, the developer stratified the IRFs by JC Stroke rehabilitation disease specific certification and compared the risk-adjusted measure scores between two groups. Stroke patients accounted for slightly less than 1 quarter of medical rehabilitation patients, it may be appropriate if the developer limited the comparison to the stroke patients.

**MP 5:** The methods for establishing validity of the self care items were varied and appropriate. The developer used several approaches to establish validity, ranging from an analysis and comparison of the data elements with other self-care assessments to a logistic regression model examining the association between

discharge self-care scores and discharge to the community. The deveoper conducted a number of rasch analyses to determine construct fit and used a 'known group differences' approach to evaluate the relationship between change in self-care scores and stroke rehabilitation certification status.

**MP 6:**A number of tests of different types of validity were done on the self-care instrument, and all were appropriate and consistent with commonly-used methods, including the use of Rasch analysis to establish the differential difficulty of items in the self-care scale.

22. **Assess the results(s) for establishing validity**

    **Submission document: Testing attachment, section 2b2.3**

    **MP 2:**All results supported validity, no concerns

    **MP 1:Overuse of Rasch measurement to test theory and validate a set of items, but results are generally supportive.**

    **MP 4:**Test results support sufficient validity of the instrument.

    **MP 3:**Face validity results were based on pre-TEP survey of TEP members and were not moderate, only 57% TEP members rates the scientific soundness as high or moderately high. Given that this was done before the meeting, it would be helpful to outline their concerns and if these concerns were resolved after the TEP.

    Content validity results were acceptable. However, the results for both data elements and scale construct validity were to be expected, not particularly convincing. Comparion of measure scores by JC Stroke certification resulted in a statistically significant difference although the difference was very small, less than 1 point. It would be helpful if the developer could demonstrate that the difference would be larger if they limited the comparison to the stroke patients.

    **MP 5:** Overall, the developer successfully demonstrated the validity of this measure through the use of several approaches, the most compelling of which were the regression analysis, rasch and known group differences analyses.

    **MP 6:**Validity at the data element level (self-care assessment instrument) was good. A panel of clinical experts provided support for face validity of the measure, and the measure scores were compared for insitutions with and without JCACHO stroke rehab certification.

23. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

    **Submission document:** Testing attachment, section 2b1.

    ☒ **Yes**

    ☐ **No**

    ☐ **Not applicable** (score-level testing was not performed)

24. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?** *NOTE that data element validation from the literature is acceptable.*

    **Submission document***: Testing attachment, section 2b1.*

    ☒ **Yes**

    ☐ **No**

    ☐ **Not applicable** (data element testing was not performed)

25. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

    ☒☐ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

    ☐☒ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has NOT been conducted)

☒☐ **Low** (NOTE: Should rate LOW if you believe that there are threats to validity and/or relevant threats to validity were not assessed OR if testing methods/results are not adequate)

☐ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the score level and the data element level is required; if not conducted, should rate as INSUFFICIENT.)

26. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**

**MP 2:**All analysis demonstrated adequate validity. No concerns noted

**MP 3:**Although the developer had conducted extensive validity tests, several test methods were somewhat weak, particularly the method used for measure score validity test. The difference between certified IRFs and not certified IRFs was significant but very small, less than 1 point, not clear if the difference is meaningful at all.

**MP 5:** The logit model provided evidence of a link between discharge self-care scores and discharge to the community (i.e., criterion validity). However, there is a likelihood that unmeasured factors in this logit model account for the relationship such as social support and the availability of community based services. At the data element level, validity analyses based on the rasch analyses demonstrate that the items in the scale are valid.

**MP 6:**The case for measure score validity depends heavility on face validity, and the developers did do a formal evaluation of face validity with an expert panel. They also noted a difference in scores between facilities with and without JCAHO stroke rehab certification. Neither of these provides strong, compelling evidence for measure score validity, but they suggest validity. Data element validity, though, is strong.

### ADDITIONAL RECOMMENDATIONS

27. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**

**MP 3:**Rasch analysis showed that 7 self-care items are of different degrees of difficulty. This implies that the difference between 10 and 20 may not be the same as the differene between 30 and 40. This has implications for across IRFs comparisons because this is a measure based on change score.

**Committee Pre-evaluation Comments:**
**Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)**

**2a1. Specifications**: *Which data elements, if any, are not clearly defined? Which codes with descriptors, if any, are not provided? Which steps, if any, in the logic or calculation algorithm or other specifications (e.g., risk/case-mix adjustment, survey/sampling instructions) are not clear? What concerns do you have about the likelihood that this measure can be consistently implemented?*

- The developer indicates the measure uses standardized data elements for the collection of functional status data. Developer also indicates that the measure does not have a simple form for the numerator or the denominator and indicates the risk adjustment formula used, which reduces the complexities associated with calculating the measure.
- No concerns about reliability
- 5. 2a1. Reliability-Specifications: Which data elements, if any, are not clearly defined? No concern here. Which codes with descriptors, if any, are not provided? No concern here. Which steps, if any, in the logic or calculation algorithm or other specifications (e.g., risk/case-mix adjustment, survey/sampling instructions) are not clear? Not evaluated or adjusted for healthcare disparities (e.g. Social Determinants of Health, socio demographic differences, etc. What concerns do you have about the likelihood that this measure can be consistently implemented? Not evaluated or adjusted for healthcare disparities (e.g. Social Determinants of Health, socio demographic differences, etc.

**2a2. Reliability testing**: *Do you have any concerns about the reliability of the measure?*

- No concerns are noted. Table 3 shows a total of 1,117 IRFs participated in data collection activities and reflected strong instrument reliability and provider-level reliability of the computed measure scores.
- No
- 6. 2a2. Reliability - Testing: Do you have any concerns about the reliability of the measure? Not at present.

**2b2. Validity testing**: *Do you have any concerns with the testing results?*

- No concerns are noted.
- No
- 7. 2b1. Validity -Testing: Do you have any concerns with the testing results? There appears to be little change in this measure over a two year period when comparing this measure with 2286, which does show significant change over a 4 year period as previously noted. This raises concerns as it relates to the discussion on harmonization of competing measures.

**Validity- Threats to Validity**: *Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data). 2b4. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? 2b5. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate they produce comparable results? 2b6. Missing data/no response: Does missing data constitute a threat to the validity of this measure?*

- These results demonstrate the ability of the measures to differentiate among facilities based on facility-level measure performance. For each IRF, the developer calculated the 95% confidence interval for the computed performance measure score and compared this with the national mean observed change score. Facilities having a confidence interval that was lower than the national mean observed change score were considered to have performed worse than the national average. Those facilities with confidence intervals that were higher than the national mean observed change score were considered to have performed better than the national average. Facilities with overlapping confidence intervals with the national mean observed change score were considered to have similar to national average performance. No concerns because missing represents less than 1%.
- (pass)
- 8. 2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data) Not evaluated or adjusted for healthcare disparities (e.g. Social Determinants of Health, socio demographic differences, etc. 2b4. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? Not sure if these are cross-correlated with longitudinal improvement in functional status of patients ultimately discharged from an IRF to home or a different post-acute setting. Since this measure is only for the Medicare population, it behooves 2b5. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate they produce comparable results? NA 2b6. Missing data/no response: Does missing data constitute a threat to the validity of this measure? As noted in the 2015 Standing Committee voting summary on page 69 of the Competing Measures Memo: The Committee noted that the measure is proposed for use for Medicare only, and felt that this limits the use of the measure and potentially introduces duplication of efforts if using multiple tools for differing payer populations

**Other Threats to Validity**: *Other Threats to Validity (Exclusions, Risk Adjustment). 2b2. Exclusions: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately excluded from the measure? 2b3. Risk Adjustment: If outcome (intermediate, health, or PRO-based) or resource use performance measure: Is there a conceptual relationship between potential social risk factor variables and the measure focus? How well do social risk factor variables that were available and analyzed align with the conceptual description provided? Are all of the risk-adjustment variables present at the start of care (if not, do you agree with the rationale provided)? Was the risk adjustment (case-mix adjustment) appropriately developed and tested? Do analyses indicate acceptable results? Is an appropriate risk-adjustment strategy included in the measure?*

- It is unclear whether the inclusion or exclusion of social risk factors in the risk adjustment model would have impacted the results. This opinion is based on the developers discussion variation of patient experiences among certain racial and ethnic groups, who had significantly different outcomes.
- No concerns

- 2b2. Exclusions: Are the exclusions consistent with the evidence? I am not clear on the purpose of this question. There are many exclusions but I am not certain these help or hinder the end-users of this measure. Are any patients or patient groups inappropriately excluded from the measure? None, although as noted in my 2b6 response, this is a Medicare only measure. 2b3. Risk Adjustment: If outcome (intermediate, health, or PRO-based) or resource use performance measure: Is there a conceptual relationship between potential social risk factor variables and the measure focus? Unclear and not well specified. How well do social risk factor variables that were available and analyzed align with the conceptual description provided? Unclear and not well specified. Are all of the risk-adjustment variables present at the start of care (if not, do you agree with the rationale provided)? No concern here. Was the risk adjustment (case-mix adjustment) appropriately developed and tested? No concern here. Do analyses indicate acceptable results? From a quantitative standpoint, no concern here. Is an appropriate risk-adjustment strategy included in the measure? No concern here.

## Criterion 3. Feasibility

**Maintenance measures – no change in emphasis – implementation issues may be more prominent**

**3. Feasibility** is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- Generated during provision of care; all data elements in defined fields in electronic health records.
- Data collected via IRF-PAI; required as part of the IRF Quality Reporting Program (QRP) and starting in October 2019, will be required by CMS for the IRF Prospective Payment System.
- Software is free and trainings were provided; no costs associated with fees, licensing, etc. Providers were given more than one year to prepare for implementation.

***Questions for the Committee:***

- Does the Committee agree with the staff assessment that there are no significant feasibility challenges associated with this measure?

**Preliminary rating for feasibility:** ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

**Committee Pre-evaluation Comments:**
**Criteria 3: Feasibility**

**3. Feasibility**: *Which of the required data elements are not routinely generated and used during care delivery? Which of the required data elements are not available in electronic form (e.g., EHR or other electronic sources)? What are your concerns about how the data collection strategy can be put into operational use?*

- The data elements are standardized and routinely generated during care delivery. No concerns are noted.
- No concerns; measure seems feasible as the information is easily available in the EHR currently.
- 10. 3. Feasibility: Which of the required data elements are not routinely generated and used during care delivery? Since the submission of these data are required for revenue cycle activities and subsequent billing to CMS, the data elements should be readily available in most/all electronic systems used in the IRF setting. Which of the required data elements are not available in electronic form (e.g., EHR or other electronic sources)? NA What are your concerns about how the data collection strategy can be put into operational use? None currently as the data structures are quite well standardized and specifications are regularly reviewed and updated.

## Criterion 4: Usability and Use

**Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact/improvement and unintended consequences**

### 4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

**4a. Use** evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

**4a.1. Accountability and Transparency.** Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

**Current uses of the measure**

**Publicly reported?** ☒ **Yes** ☐ **No**

**Current use in an accountability program?** ☒ **Yes** ☐ **No** ☐ **UNCLEAR**

**OR**

**Planned use in an accountability program?** ☐ **Yes** ☐ **No**

**Accountability program details**

- IRF QRP
- IRF Compare (collecting 2019 data, will be reported in 2020)

**4a.2. Feedback on the measure by those being measured or others.** Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

**Feedback on the measure by those being measured or others**

- Providers receive results and assistance with interpretation via confidential feedback reports, provider training seminars, manuals and materials, and responses to questions submitted to the IRF QRP Help Desk and IRF Public Reporting Help Desk.
- Patients and families and other stakeholders can review results on the publicly available IRF Compare.
- In the 2015 and 2019 rule proposals, public commenters mostly supported the addition of this measure to the IRF QRP. The developer also gathered feedback from a TEP in 2017, and some members expressed support. The developer has made updates to the measure based on stakeholder feedback including to the risk adjustment model.

**Additional Feedback:** N/A

***Questions for the Committee:***

- How can the performance results be used to further the goal of high-quality, efficient healthcare?
- How has the measure been vetted in real-world settings by those being measured or others?

**Preliminary rating for Use:** ☒ **Pass** ☐ **No Pass**

### 4b. Usability (4a1. Improvement; 4a2. Benefits of measure)

**4b. Usability** evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

**4b.1 Improvement.** Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

**Improvement results**

- Because this measure is very recently implemented and has not yet been reported, there is no trend in performance over time data. The measure has remained stable over FY 2017 and calendar year 2017.

**4b2. Benefits vs. harms.** Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

**Unexpected findings (positive or negative) during implementation**

- None found

**Potential harms**

- None found

**Additional Feedback:** N/A

***Questions for the Committee:***

- How can the performance results be used to further the goal of high-quality, efficient healthcare?
- Do the benefits of the measure outweigh any potential unintended consequences?

**Preliminary rating for Usability and use:** ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

**Committee Pre-evaluation Comments:**
**Criteria 4: Usability and Use**

**4a. Use**: *4a1. Use - Accountability and Transparency: How is the measure being publicly reported? Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? For maintenance measures - which accountability applications is the measure being used for? For new measures - if not in use at the time of initial endorsement, is a credible plan for implementation provided? 4a2. Use - Feedback on the measure: Have those being measured been given performance results or data, as well as assistance with interpreting the measure results and data? Have those being measured or other users been given an opportunity to provide feedback on the measure performance or implementation? Has this feedback has been considered when changes are incorporated into the measure?*

- Quality measure data collected in calendar year 2019 will be publicly reported in 2020 on CMS's IRF Compare website at: https://www.medicare.gov/inpatientrehabilitationfacilitycompare/. Since 2016, CMS has publicly reported IRF QRP quality measure data on the IRF Compare website. This website reports quality data for each IRF, and these data are also publicly available for download at: https://data.medicare.gov/data/inpatient-rehabilitation-facility-compare. 2. Providers receive results and assistance with interpretation via confidential feedback reports, provider training seminars, manuals and materials, and responses to questions submitted to the IRF QRP Help Desk and IRF Public Reporting Help Desk. 3. Patients and families and other stakeholders can review results on the publicly available IRF Compare. 4. In the 2015 and 2019 rule proposals, public commenters mostly supported the addition of this measure to the IRF Quality Reporting Program. The developer also gathered feedback from a TEP
- Results are publicly reported and used for accountability. Feedback has been obtained from appropriate users.
- 11. 4a1. Use - Accountability and Transparency: How is the measure being publicly reported? By CMS for comparative purposes via Inpatient Rehabilitation Facility (IRF) Compare. https://www.medicare.gov/inpatientrehabilitationfacilitycompare/ Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? Yes For maintenance measures - which accountability applications is the measure being used for? Public reporting and payment For new measures - if not in use at the time of initial endorsement, is a credible plan for

implementation provided? NA 11. 4a1. Use - Accountability and Transparency: How is the measure being publicly reported? By CMS for comparative purposes via Inpatient Rehabilitation Facility (IRF) Compare. https://www.medicare.gov/inpatientrehabilitationfacilitycompare/ Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? Yes For maintenance measures - which accountability applications is the measure being used for? Public reporting and payment For new measures - if not in use at the time of initial endorsement, is a credible plan for implementation provided? NA

**4b. Usability**: *4b1. Usability – Improvement: How can the performance results be used to further the goal of high-quality, efficient healthcare? If not in use for performance improvement at the time of initial endorsement, is a credible rationale provided that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations? 4b2. Usability – Benefits vs. harms: Describe any actual unintended consequences and note how you think the benefits of the measure outweigh them.*

- Because this measure is very recently implemented and has not yet been reported, there is no trend in performance over time data. The measure has remained stable over FY 2017 and calendar year 2017. No unexpected or potential harms found.
- Measure has potential to improve the quality of care. Benefits outweigh harms.
- 12. 4b1. Usability – Improvement: How can the performance results be used to further the goal of high-quality, efficient healthcare? There appears to be little change in 2633 measure over a two year period when compared with this measure (2286), which does show significant change over a 4 year period as previously noted. This raises concerns as it relates to the discussion on harmonization of competing measures. Since there has not been quite a significant change in 2633 over time, this measure does not currently appear to be effective as a quality improvement resource. If not in use for performance improvement at the time of initial endorsement, is a credible rationale provided that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations? Unclear 4b2. Usability – Benefits vs. harms: Describe any actual unintended consequences and note how you think the benefits of the measure outweigh them. There appears to be little change in 2633 measure over a two year period when compared with this measure (2286), which does show significant change over a 4 year period as previously noted. This raises concerns as it relates to the discussion on harmonization of competing measures. Since there has not been quite a significant change in 2633 over time, this measure appears to be an ineffective resource for achieving measurable quality improvements over time. 12. 4b1. Usability – Improvement: How can the performance results be used to further the goal of high-quality, efficient healthcare? There appears to be little change in 2633 measure over a two year period when compared with this measure (2286), which does show significant change over a 4 year period as previously noted. This raises concerns as it relates to the discussion on harmonization of competing measures. Since there has not been quite a significant change in 2633 over time, this measure appears to be an ineffective resource for achieving measurable quality improvements over time. If not in use for performance improvement at the time of initial endorsement, is a credible rationale provided that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations? NA 4b2. Usability – Benefits vs. harms: Describe any actual unintended consequences and note how you think the benefits of the measure outweigh them. Since there has not been quite a significant change in 2633 over time, this measure appears to be an ineffective resource for achieving measurable quality improvements over time.

## Criterion 5: Related and Competing Measures

**Related or competing measures**

- This measure is competing with one measure: 2286: Functional Change: Change in Self Care Score
- The Committee will need to compare both measures and attempt to reach a best-in-class decision. NQF staff will prepare additional materials to assist the Committee in this comparison.

**This measure is related to, but not competing with, the following measures:**

- 0174 : Improvement in bathing
- 0175 : Improvement in bed transferring
- 0426 : Functional status change for patients with Shoulder impairments
- 0427 : Functional status change for patients with elbow, wrist and hand impairments
- 0428 : Functional status change for patients with General orthopaedic impairments
- 0688 : Percent of Residents Whose Need for Help with Activities of Daily Living Has Increased (long stay)
- 2287 : Functional Change: Change in Motor Score
- 2613 : CARE: Improvement in Self Care
- 2643 : Average change in functional status following lumbar spine fusion surgery
- 2769 : Functional Change: Change in Self Care Score for Skilled Nursing Facilities
- 2775 : Functional Change: Change in Motor Score for Skilled Nursing Facilities
- 2776 : Functional Change: Change in Motor Score in Long Term Acute Care Facilities
- 2777 : Functional Change: Change in Self Care Score for Long Term Acute Care Facilities

**Harmonization**

- This measure is not fully harmonized with the related or competing measures.

**Committee Pre-evaluation Comments: Criterion 5:**
**Related and Competing Measures**

**Related and Competing**: *Are there any related and competing measures? If so, are any specifications that are not harmonized? Are there any additional steps needed for the measures to be harmonized?*

- Related or competing measures
- This measure is competing with one measure: 2286: Functional Change: Change in Self Care Score
- The Committee will need to compare both measures and attempt to reach a best-in-class decision. NQF staff will prepare additional materials to assist the Committee in this comparison. This measure is related to, but not competing with, the following measures:
- 0174 : Improvement in bathing
- 0175 : Improvement in bed transferring
- 0426 : Functional status change for patients with Shoulder impairments
- 0427 : Functional status change for patients with elbow, wrist and hand impairments
- 0428 : Functional status change for patients with General orthopedic impairments
- 0688 : Percent of Residents Whose Need for Help with Activities of Daily Living Has Increased (long stay)
- 2287 : Functional Change: Change in Motor Score
- 2613 : CARE: Improvement in Self Care
- 2643 : Average change in functional status following lumbar spine fusion surgery
- 2769 : Functional Change: Change in Self Care Score for Skilled Nursing Facilities
- 2775 : Functional Change: Change in Motor Score for Skilled Nursing Facilities
- 2776 : Functional Change: Change in Motor Score in Long Term Acute Care Facilities
- 2777 : Functional Change: Change in Self Care Score for Long Term Acute Care Facilities Harmonization
- This measure is not fully harmonized with the related or competing measures.
- One competing measure (2286), several related measures are noted.
- 13. 5. Related and Competing: Are there any related and competing measures? 2286 If so, are any specifications that are not harmonized? Yes. 2633 for Medicare only. Are there any additional steps

needed for the measures to be harmonized? Yes but unclear what these could be since CMS has now retired 2286.

## Public and Member Comments

**Comments and Member Support/Non-Support Submitted as of: June/13/2019**

- **No NQF members have submitted support/non-support choices as of this date**

## 1. Evidence and Performance Gap – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

**1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form**

2633_NQF_evidence_4-22-19.docx

**1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?**

Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

Yes

### 1a. Evidence (subcriterion 1a)

**Measure Number** (*if previously endorsed*)**:** NQF #2633

**Measure Title**: Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Change in Self-Care Score for Medical Rehabilitation Patients

**IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:** Not applicable

**Date of Submission**: 4/9/2019

**1a.1.This is a measure of**: (*should be consistent with type of measure entered in De.1*)

Outcome

☒ Outcome: Change in Function: Self-Care

☐Patient-reported outcome (PRO):

*PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors.* (*A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)*

☐ Intermediate clinical outcome (*e.g., lab value*):

☐ Process:

☐ Appropriate use measure:

☐ Structure:

☐ Composite:

**1a.2 LOGIC MODEL** Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

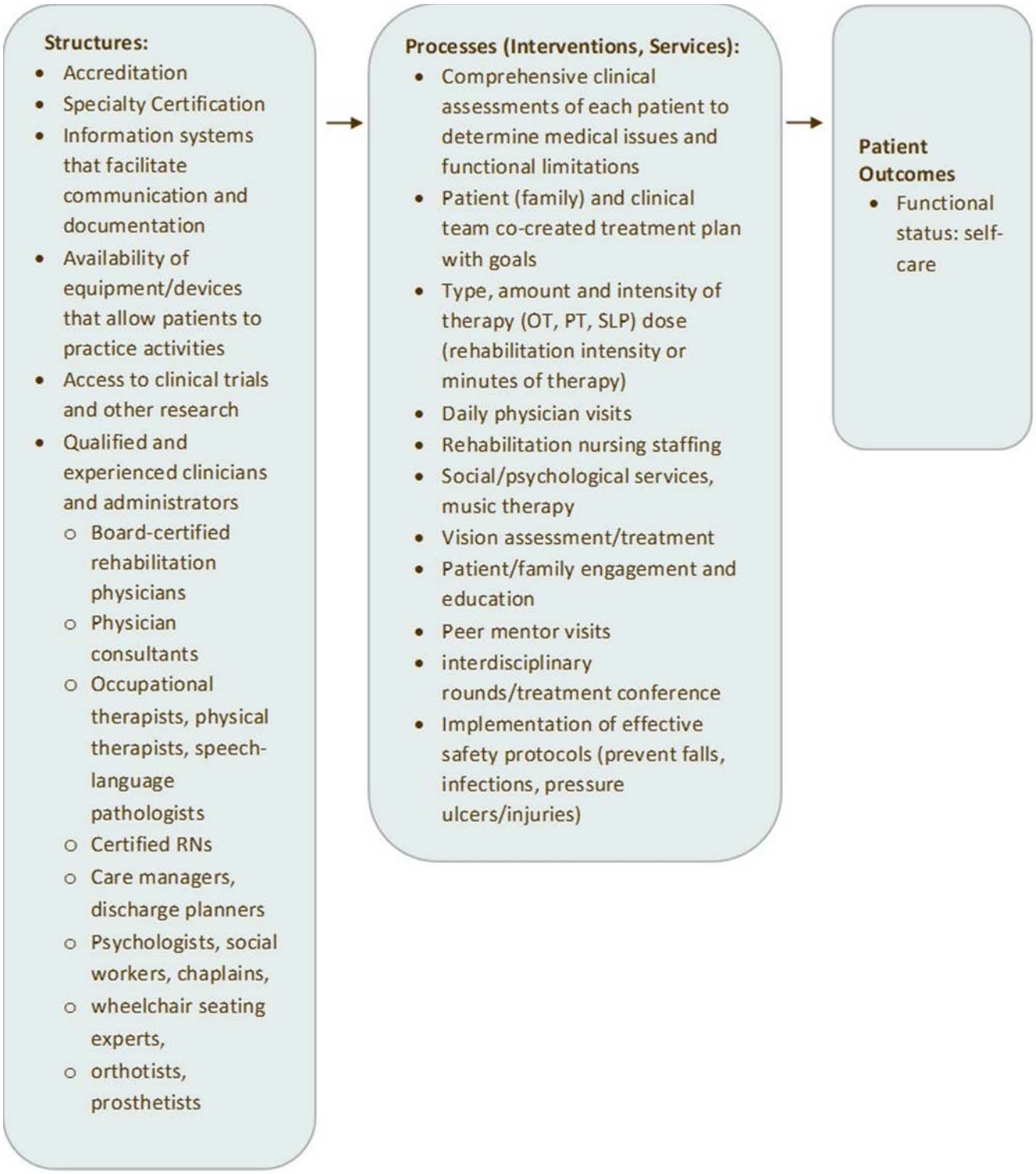
Inpatient Rehabilitation Facilities (IRFs) are designed to provide intensive rehabilitation services to patients. Patients seeking care in IRFs are those whose illness, injury, or condition has resulted in a loss of function, and

for whom rehabilitative care is expected to help regain that function. Examples of conditions treated in IRFs include stroke, spinal cord injury, hip fracture, brain injury, neurological disorders, and other diagnoses characterized by loss of function. During an IRF stay, goals of treatment include fostering the patient's ability to manage his or her daily activities so that the patient can complete self-care and mobility activities as independently as possible and, if feasible, return to a safe, active, and productive life in a community-based setting.

**NQF Evidence 2019**

Key rehabilitation services provided to patients in an IRF include physiatry care, (i.e., physical medicine and rehabilitation physician), physical therapy, occupational therapy, rehabilitation nursing, speech-language pathology, and prosthetic and orthotic services (Medicare Payment Advisory Commission, 2019). Figure 1a lists the structures, processes and the outcomes that relate to this measure. This model shows that an IRF's structures and processes (treatments or interventions) can result in improved patient functioning.

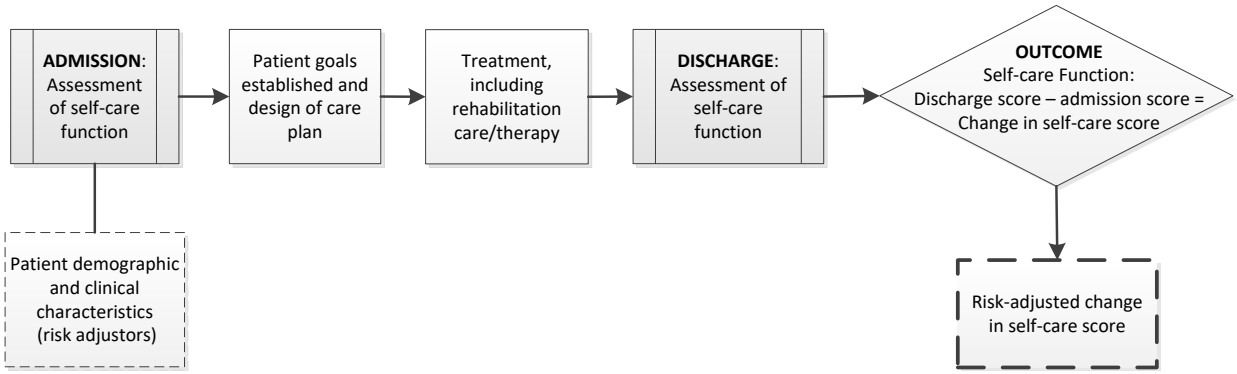
Figure 1a Structures and Processes Associated with Patients' Functional Outcomes.



**From previous NQF Submission (2014)**

Given that the primary goal of rehabilitation is improvement in function, IRF clinicians have traditionally assessed and documented patients' functional status at admission and discharge to evaluate the effectiveness of the rehabilitation care provided to individual patients, as well as the effectiveness of the IRF overall (see **Figure 1**).

**Figure 1. Role of Patient Assessment, Interventions and Functional Outcomes (2014 NQF Submission)**




**1a.3 Value and Meaningfulness:** **IF** this measure is derived from patient report, provide evidence that the target population values the measured ***outcome, process, or structure*** and finds it meaningful. (Describe how and from whom their input was obtained.)

Not applicable. This measure uses data that is clinician-reported.

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

**1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.**

To demonstrate that IRFs have the ability to improve patient functioning, including self-care abilities, NQF requires evidence that at least one structure, process, intervention or service can affect patient functioning. Because intensive, interdisciplinary therapy services are a core feature of an IRF stay and these services are targeted to improve functional outcomes, we provide a summary of evidence from the literature that is focused on therapy services and functional outcomes.

For this evidence update, we conducted a scoping review to identify relevant literature examining the relation between therapy interventions and improved patient functioning. We describe the details about the scoping review methodology below, after the literature summary and abstracts. However, we would like to note upfront that the summary below only includes articles published after January 1, 2013 since we sought to identify relevant literature since our 2014 NQF submission.

**Therapy Interventions and Functional Outcomes**

Most IRF research examining functional outcomes has focused on motor function, which encompasses self-care and mobility and sometimes bladder function. Several observational studies have reported positive associations between the amount of therapy provided and motor function for patients with various diagnoses, including spinal cord injury (Backus et al., 2013), stroke (Wang et al., 2013; Goedert, Zhang & Barrett, 2015), traumatic brain injury (Rosenbaum Gordon, Joannou, & Berman, 2018), hip fracture (Siebens et al., 2013); and after West Nile Virus (Hoffman & Paschal, 2013). One additional study, that was not diagnosis specific, also found improved functional outcomes related to rehabilitation therapy intensity (Morghen et al., 2017). Backus (2013) found that more time in inpatient physical therapy (PT) was associated with higher motor function 1-year post-discharge, while Wang (2013) reported a significant relationship between daily therapeutic duration and functional gain during an IRF stay and offered treatment time thresholds for optimal functional outcomes for patients with stroke.

Two additional observational studies examined the influence of age as a mediator on the amount of therapy provided and patients' functional outcomes. One study found that older adults (65 and older) with a traumatic brain injury received fewer hours of treatment per day and fewer total hours of therapy due to a shorter length of stay, and these patients overall regained less function, compared to younger IRF patients (Djikers et al., 2013). Hsieh et. al, (2013) found that older adults (60 and older) with a spinal cord injury received less rehabilitation therapy during longer rehabilitation stays, and had lower functional abilities compared to younger patients.

Several observational studies reported that IRF care improved patients' motor functional outcomes but did not specify the type and amount of therapy provided. Improvement in motor functional outcomes were reported for patients recovering from trauma (Hamidi et al., 2018), patients 85 and older post-stroke (O'Brien & Xue, 2016), post-knee surgery patients (Chu et al., 2016), and post-hip fracture patients (Cary, Baernholdt, Anderson, & Merwin, 2015). Hamidi et al. (2018) also explored the relationship between frailty level and functional improvement among IRF patients, finding that frailer patients were less likely to regain their baseline functional ability.

Studies examining specific rehabilitation therapy interventions and patients' functional outcomes has generally been challenging to examine (Kroll & Fisher, 2018), because rehabilitation interventions tend to be multidisciplinary, tailored to each patient's specific needs and there are no standardized definitions and no standardized measurement of interventions. Efforts are underway to classify interventions using standardized terminology in order to better understand the relation between interventions and outcomes; that is, the active ingredients of a rehabilitation program. Winstein (2016) noted that measuring the effect of rehabilitation interventions in post-acute rehabilitation settings is important for understanding "the amount of adequate resources, dose and duration," that are needed to affect functional improvement.

We identified two IRF studies that examined the effect of novel therapy interventions that were added to "usual" therapy. Herron (2016) found that occupational therapists conducting initial visual assessments of stroke patients resulted in better patient functional outcomes. Another study found that adding attention-control training for stroke patients improved motor outcomes compared to control participants over a 6-month period (Skidmore et al., 2015).

Rehabilitation therapy services are provided in all types of post-acute care services: IRFs, skilled nursing facilities, home health and long-term care hospitals. IRFs have historically provided the most intensive therapy services, and several studies have compared patients' functional outcomes by type of post-acute care setting to better understand the role of therapy intensity and other IRF-specific care. Three studies reported improved motor functional outcomes for IRF patients compared to other post-acute care settings. Sauter (2013) found that patients who underwent major dysvascular lower limb amputations receipt of interdisciplinary rehabilitation services in an IRF yielded improved functional outcomes 6 months after amputation compared to care received in SNFs or at home. Nehra et al. (2016) reported improved motor function for post-trauma patients discharged to an IRF compared to those who were not treated in an IRF. A systematic review exploring poststroke outcomes in IRFs compared to SNFs, found that of four included studies that compared functional outcomes, one study reported higher functional ratings for SNFs compared to IRFs, three studies found functional gains were larger for IRF patients compared to patients treated in skilled nursing facilities (Alcusky, Ulbricht, & Lapane, 2018). One of the studies included in the systematic review. Chan et al. (2013) found that stroke patients who received therapy in an IRF had higher functional motor gains compared to patient treated in skilled nursing facilities and home health.

In addition to the studies that compare functional outcome across post-acute care settings, we provide a summary of studies examining therapy services and functional outcomes in the SNF setting. These studies are pertinent, because there is overlap in the types of patients treated in IRFs and SNFs, and the amount of therapy provided in SNFs tends to vary more than in the IRF setting.

Positive associations of functional motor outcomes for patients receiving therapy in SNFs have been reported for several observation studies. One SNF study found that more than 60 percent of patients improved their

functional status as a result of their SNF stay, and noted that patients with conditions such as cognitive impairment, delirium, dementia, heart failure, and stroke showed less improvement in daily activity performance during their stay. Likewise, among traumatic brain injurt patients, SNF patients admitted with cognitive or communication impairments gained function, but tended to have less improvement in motor function (Wysocki, Thomas, & Mor, 2015). Jung (2016) examined temporal trends in therapy provision in SNFs and found therapy hours increased 52% between 2000 and 2009, and that more therapy hours in SNFs appeared to improve outcomes, except for patients who were already receiving a high level of therapy.

**Conclusion**

In summary, as required by NQF's endorsement maintenance process, we sought to identify evidence linking a healthcare structure or process (interventions or services) to patient outcomes. For this review, we summarized various studies pertaining to the relation between therapy interventions and services and IRF patients' functional outcomes. We also included peer-reviewed evidence from other post-acute care (PAC) settings. Although no identified articles focused on self-care outcomes specifically, the majority included motor function which encompasses self-care, mobility, and sometimes bladder functioning. Most articles were observational studies while two presented novel therapies being tested in addition to "usual" therapy. One article focused on rehabilitation guidelines for PAC settings. This review provides supportive evidence that functional improvement in IRF patients is related to the therapy interventions they received while in an IRF.

**Citations and Abstracts**

Alcusky, M., Ulbricht, C. M., & Lapane, K. L. (2018). Postacute Care Setting, Facility Characteristics, and Poststroke Outcomes: A Systematic Review. *Arch Phys Med Rehabil, 99*(6), 1124-1140.e1129. doi:10.1016/j.apmr.2017.09.005

OBJECTIVES: To synthesize research comparing poststroke health outcomes between patients rehabilitated in skilled nursing facilities (SNFs) and those in inpatient rehabilitation facilities (IRFs) as well as to evaluate relations between facility characteristics and outcomes.

DATA SOURCES: PubMed and CINAHL searches spanned January 1, 1998, to October 6, 2016, and encompassed MeSH and free-text keywords for stroke, IRF/SNF, and study outcomes. Searches were restricted to peer-reviewed research in humans published in English.

STUDY SELECTION: Observational and experimental studies examining outcomes of adult patients with stroke rehabilitated in an IRF or SNF were eligible. Studies had to provide site of care comparisons and/or analyses incorporating facility-level characteristics and had to report >/=1 primary outcome (discharge setting, functional status, readmission, quality of life, all-cause mortality). Unpublished, single-center, descriptive, and non-US studies were excluded. Articles were reviewed by 1 author, and when uncertain, discussion with study coauthors achieved consensus. Fourteen titles (0.3%) were included.

DATA EXTRACTION: The types of data, time period, size, design, and primary outcomes were extracted. We also extracted 2 secondary outcomes (length of IRF/SNF stay, cost) when reported by included studies. Effect measures, modeling approaches, methods for confounding adjustment, and potential confounders were extracted. Data were abstracted by 1 author, and the accuracy was verified by a second reviewer.

DATA SYNTHESIS: Two studies evaluating community discharge, 1 study evaluating the predicted probability of readmission, and 3 studies evaluating all-cause mortality favored IRFs over SNFs. Functional status comparisons were inconsistent. No studies evaluated quality of life. Two studies confirmed increased costs in the IRF versus SNF setting. Although substantial facility variation was described, few studies characterized sources of variation.

CONCLUSIONS: The few studies comparing poststroke outcomes indicated better outcomes (with higher costs) for patients in IRFs versus those in SNFs. Contemporary research on the role of the

postacute care setting and its attributes in determining health outcomes should be prioritized to inform reimbursement system reform.

Backus, D., Gassaway, J., Smout, R. J., Hsieh, C. H., Heinemann, A. W., DeJong, G., & Horn, S. D. (2013). Relation between inpatient and postdischarge services and outcomes 1 year postinjury in people with traumatic spinal cord injury. *Arch Phys Med Rehabil, 94*(4 Suppl), S165-174. doi:10.1016/j.apmr.2013.01.012

OBJECTIVE: To examine the association between inpatient and postdischarge rehabilitation services and function, life satisfaction, and community participation 1 year after spinal cord injury (SCI).
DESIGN: Prospective, observational.

SETTING: Six rehabilitation facilities.

PARTICIPANTS: Patients with SCI (N=1376).

INTERVENTIONS: None.

MAIN OUTCOME MEASURES: Satisfaction with Life Scale (SWLS), Craig Handicap Assessment and Reporting Technique (CHART), motor FIM (mFIM), and return to work/school at 1 year post-SCI.
RESULTS: Demographic and injury characteristics explained 49% of the variance in mFIM and 9% to 25% of the variance in SWLS and CHART social integration, mobility, and occupation scores. Inpatient rehabilitation services explained an additional 2% of the variance for mFIM and 1% to 3% of the variance for SWLS and CHART scores. More time in inpatient physical therapy (PT) was associated with higher mFIM scores; more time in inpatient therapeutic recreation (TR) and social work and more postdischarge nursing (NSG) were associated with lower mFIM scores. More inpatient PT and TR and more postdischarge PT were associated with higher mobility scores; more inpatient psychology (PSY) was associated with lower mobility scores. More postdischarge TR was associated with higher SWLS; more postdischarge PSY services was associated with lower SWLS. Inpatient TR was positively associated with social integration scores; postdischarge PSY was negatively associated with social integration scores. More postdischarge vocational counseling was associated with higher occupation scores. Differences between centers did not explain additional variability in the outcomes studied.

CONCLUSIONS: Inpatient and postdischarge rehabilitation services are weakly associated with life satisfaction and societal participation 1 year after SCI. Further study of the type and intensity of postdischarge services, and the association with outcomes, is needed to ascertain the most effective use of therapy services after SCI.

Cary, M. P., Baernholdt, M., Anderson, R. A., & Merwin, E. I. (2015). Performance-based outcomes of inpatient rehabilitation facilities treating hip fracture patients in the United States. *Arch Phys Med Rehabil, 96*(5), 790-798. doi:10.1016/j.apmr.2015.01.003

OBJECTIVE: To examine the influence of facility and aggregate patient characteristics of inpatient rehabilitation facilities (IRFs) on performance-based rehabilitation outcomes in a national sample of IRFs treating Medicare beneficiaries with hip fracture.

DESIGN: Secondary data analysis.

SETTING: U.S. Medicare-certified IRFs (N=983).

PARTICIPANTS: Data included patient records of Medicare beneficiaries (N=34,364) admitted in 2009 for rehabilitation after hip fracture.

INTERVENTION: Not applicable.

MAIN OUTCOME MEASURES: Performance-based outcomes included mean motor function on discharge, mean motor change (mean motor score on discharge minus mean motor score on admission), and percentage discharged to the community.

RESULTS: Higher mean motor function on discharge was explained by aggregate characteristics of patients with hip fracture (lower age [P=.009], lower percentage of blacks [P<.001] and Hispanics [P<.001], higher percentage of women [P=.030], higher motor function on admission [P<.001], longer

length of stay [P<.001]) and facility characteristics (freestanding [P<.001], rural [P<.001], for profit [P=.048], smaller IRFs [P=.014]). The findings were similar for motor change, but motor change was also associated with lower mean cognitive function on admission (P=.008). Higher percentage discharged to the community was associated with aggregate patient characteristics (lower age [P<.001], lower percentage of Hispanics [P=.009], higher percentage of patients living with others [P<.001], higher motor function on admission [P<.001]). No facility characteristics were associated with the percentage discharged to the community. CONCLUSIONS: Performance-based measurement offers health policymakers, administrators, clinicians, and consumers a major opportunity for securing health system improvement by benchmarking or comparing their outcomes with those of other similar facilities. These results might serve as the basis for benchmarking and quality-based reimbursement to IRFs for 1 impairment group: hip fracture.

Chan, L., Sandel, M. E., Jette, A. M., Appelman, J., Brandt, D. E., Cheng, P., . . . Rasch, E. K. (2013). Does postacute care site matter? A longitudinal study assessing functional recovery after a stroke. *Arch Phys Med Rehabil, 94*(4), 622-629. doi:10.1016/j.apmr.2012.09.033

OBJECTIVE: To determine the impact of postacute care site on stroke outcomes. DESIGN: Prospective cohort study. SETTING: Four northern California hospitals that are part of a single health maintenance organization. PARTICIPANTS: Patients with stroke (N=222) enrolled between February 2008 and July 2010. INTERVENTION: Not applicable. MAIN OUTCOME MEASURE: Baseline and 6-month assessments were performed using the Activity Measure for Post Acute Care (AM-PAC), a test of self-reported function in 3 domains: Basic Mobility, Daily Activities, and Applied Cognition. RESULTS: Of the 222 patients analyzed, 36% went home with no treatment, 22% received home health/outpatient care, 30% included an inpatient rehabilitation facility (IRF) in their care trajectory, and 13% included a skilled nursing facility (but not IRF) in their care trajectory. At 6 months, after controlling for important variables such as age, functional status at acute care discharge, and total hours of rehabilitation, patients who went to an IRF had functional scores that were at least 8 points higher (twice the minimally detectable change for the AM-PAC) than those who went to a skilled nursing facility in all 3 domains and in 2 of 3 functional domains compared with those who received home health/outpatient care. CONCLUSIONS: Patients with stroke may make more functional gains if their postacute care includes an IRF. This finding may have important implications as postacute care delivery is reshaped through health care reform.

Chu, S. K., Babu, A. N., McCormick, Z., Mathews, A., Toledo, S., & Oswald, M. (2016). Outcomes of Inpatient Rehabilitation in Patients With Simultaneous Bilateral Total Knee Arthroplasty. *Pm r, 8*(8), 761-766. doi:10.1016/j.pmrj.2015.11.005

BACKGROUND: The number of total knee arthroplasty (TKA) procedures performed in the United States is increasing each year, and the number of bilateral TKA procedures has also increased during the past 2 decades. However, few studies in the literature have investigated the rehabilitation outcomes of patients who undergo bilateral TKA. This study was performed to provide information on the benefits and role of inpatient rehabilitation for patients after bilateral TKA. OBJECTIVE: To investigate the functional outcomes, complications, and transfer rates of patients in the inpatient rehabilitation setting who undergo simultaneous bilateral TKA. DESIGN: Retrospective cohort study. SETTING: Freestanding inpatient rehabilitation hospital. PATIENTS: Ninety-four patients admitted to an inpatient rehabilitation hospital after simultaneous bilateral TKA from 2008-2013. METHODS: Retrospective chart review of demographic, clinical, and functional data for patients admitted to inpatient rehabilitation after simultaneous bilateral TKA. MAIN OUTCOME MEASURES: Length of stay, admission and discharge Functional Independence Measure (FIM), and FIM efficiency. RESULTS: The study included 27 male (28.7%) and 67 female (71.3%) patients aged 42.0-86.9 years, with a mean of 65.6 +/- 10.2 years. Mean length of time between surgery and admission to inpatient rehabilitation was 4.5 +/- 3.3 days. Mean length of stay in rehabilitation was 11.7 +/- 4.2 days. Mean admission and discharge FIM scores were 87.3 +/- 11.7 and 113.4 +/- 4.8, respectively, with a mean FIM gain of 26.1

+/- 10.5. The mean FIM efficiency was 2.33 +/- 0.84. Eight patients required transfer to an acute care hospital. Complications leading to transfer to acute care facilities included sepsis, cardiac arrhythmias, knee dislocation, and suspected small bowel obstruction. Eighty-eight patients were discharged home, 4 patients were discharged to skilled nursing facilities, and 2 patients were transferred to an acute care hospital and did not return to the inpatient rehabilitation hospital. CONCLUSIONS: After undergoing simultaneous bilateral TKA, patients demonstrate functional gains when admitted to inpatient rehabilitation facilities based on FIM gains and FIM efficiency scores; 8.5% of patients in this cohort required transfer to an acute care facility as a result of complications during inpatient rehabilitation, and 93.6% of patients were discharged home.

Dijkers, M., Brandstater, M., Horn, S., Ryser, D., & Barrett, R. (2013). Inpatient rehabilitation for traumatic brain injury: the influence of age on treatments and outcomes. *NeuroRehabilitation, 32*(2), 233-252. doi:10.3233/nre-130841

BACKGROUND: Elderly persons with traumatic brain injury (TBI) are increasingly admitted to inpatient rehabilitation, but we have limited knowledge of their characteristics, the treatments they receive, and their short-term and medium-term outcomes. This study explored these issues by means of comparisons between age groups. METHODS: Data on 1419 patients admitted to 9 inpatient rehabilitation facilities for initial rehabilitation after TBI were collected by means of (1) abstraction from medical records; (2) point-of care forms completed by therapists after each treatment session; and (3) interviews at 3 months and 9 months after discharge, conducted with the patient or a proxy. RESULTS: Elderly persons (65 or older) had a lower brain injury severity, and a shorter length of stay (LOS) in acute care. During rehabilitation, they received fewer hours of therapy, due to a shorter LOS and fewer hours of treatment per day, especially from psychology and therapeutic recreation. They regained less functional ability during and after inpatient rehabilitation, and had a very high mortality rate. CONCLUSIONS: Elderly people can be rehabilitated successfully, and discharged back to the community. The treatment therapists deliver, and issues surrounding high mortality need further research.

Goedert, K. M., Zhang, J. Y., & Barrett, A. M. (2015). Prism adaptation and spatial neglect: the need for dose-finding studies. Front Hum Neurosci, 9, 243. doi:10.3389/fnhum.2015.00243

Spatial neglect is a devastating disorder in 50-70% of right-brain stroke survivors, who have problems attending to, or making movements towards, left-sided stimuli, and experience a high risk of chronic dependence. Prism adaptation is a promising treatment for neglect that involves brief, daily visuo-motor training sessions while wearing optical prisms. Its benefits extend to functional behaviors such as dressing, with effects lasting 6 months or longer. Because one to two sessions of prism adaptation induce adaptive changes in both spatial-motor behavior (Fortis et al., 2011) and brain function (Saj et al., 2013), it is possible stroke patients may benefit from treatment periods shorter than the standard, intensive protocol of ten sessions over two weeks-a protocol that is impractical for either US inpatient or outpatient rehabilitation. Demonstrating the effectiveness of a lower dose will maximize the availability of neglect treatment. We present preliminary data suggesting that four to six sessions of prism treatment may induce a large treatment effect, maintained three to four weeks post-treatment. We call for a systematic, randomized clinical trial to establish the minimal effective dose suitable for stroke intervention.

Hamidi, M., Zeeshan, M., O'Keeffe, T., Nisbet, B., Northcutt, A., Nikolich-Zugich, J., . . . Joseph, B. (2018). Prospective evaluation of frailty and functional independence in older adult trauma patients. *Am J Surg, 216*(6), 1070-1075. doi:10.1016/j.amjsurg.2018.10.023

BACKGROUND: The aim of our study was to assess the association between frailty and functional status in geriatric trauma patients. METHODS: 3-year(2013-2015) prospective analysis and included all geriatric trauma patients(>/=65y) discharged to a single rehabilitation center from our level-I trauma center. Frailty was measured using Trauma-Specific-Frailty-Index(TSFI) while Functional status was

assessed using functional-independence-measure(FIM) at admission and discharge from rehabilitation center. Multivariate linear regression analysis was performed. RESULTS: 267 patients were enrolled. Mean age was 76.9+/-7.1y, 63.6% were males. Overall, 22.8% were frail, and 37.4% were pre-frail. On linear regression, higher motor-FIM, higher cognitive-FIM scores at admission, and longer length-of-stay at rehab were independently associated with increased discharge FIM score. While, ISS(injury-severity-score), pre-frail and frail status were negatively correlated with FIM gain. CONCLUSION: Frail patients were less likely to recover to their baseline functional status compared with non-frail patients. Early focused intervention in frail elderly patients is warranted to improve functional status in this population.

Herron, S. (2016). Review of experience with a collaborative eye care clinic in inpatient stroke rehabilitation. *Top Stroke Rehabil, 23*(1), 67-75. doi:10.1179/1074935715z.00000000065

BACKGROUND: Visual deficits following stroke are frequently subtle and are often overlooked. Even though these visual deficits may be less overt in nature, they are still debilitating to survivors. Visual deficits have been shown to negatively impact cognition, mobility, and activities of daily living (ADL). There is little consistency across healthcare facilities regarding protocol for assessing vision following stroke. OBJECTIVE: This research was designed to describe a profile for patients exhibiting visual deficits following stroke, examine the role of occupational therapists in vision assessment, and discuss a potential model to provide a protocol for collaboration with an eye care professional as part of the rehabilitation team. METHODS: The sample consisted of 131 patients in an inpatient rehabilitation (IPR) unit who were identified as having potential visual deficits. Occupational therapists on an IPR unit administered initial vision screenings and these patients were subsequently evaluated by the consulting optometrist. Frequencies were calculated for the appearance of functional symptoms, diagnoses, and recommendations. Correlations were also computed relating diagnoses and recommendations made. RESULTS: All patients referred by the occupational therapist for optometrist evaluation had at least one visual diagnosis. The most frequent visual diagnoses included: saccades (77.7%), pursuits (61.8%), and convergence (63.4%). There was also a positive correlation between number of functional symptoms seen by occupational therapists and visual diagnoses made by the optometrist ($r = 0.209$, $P = 0.016$). CONCLUSION: Results of this study support the need for vision assessment following stroke in IPR, confirm the role of occupational therapists in vision assessment, and support the need for an optometrist as a member of the rehabilitation team.

Hoffman, J. E., & Paschal, K. A. (2013). Functional outcomes of adult patients with West Nile virus admitted to a rehabilitation hospital. *J Geriatr Phys Ther, 36*(2), 55-62. doi:10.1519/JPT.0b013e318258bcba

BACKGROUND AND PURPOSE: The clinical manifestation of West Nile Virus (WNV) varies in individuals from mild flu-like symptoms to acute flaccid paralysis. Advanced age is the most significant risk factor for developing severe neurological disease and for death. The broad range of neurologic symptoms associated with WNV infection leads to varied body structure and function limitations and participation restrictions that may require rehabilitation. The purpose of this study is to describe the functional impairments upon admission and the functional outcomes at discharge of 48 adult patients admitted with WNV to a rehabilitation facility in the Midwest from 2002 to 2009. METHODS: A retrospective chart review was completed on 48 patients (29 male, 19 female) with mean age 67.8 (SD = 16.6, range = 24-91) years and median age 72.5 years, admitted to inpatient rehabilitation with a diagnosis of WNV after January 1, 2002, and discharged prior to December 31, 2009. General information (sex, age, social history, employment, and living environment), past medical history, and information specific to the current hospitalization (medical conditions, functional status and activity level on admission and discharge as measured by the Functional Independence Measure [FIM], lengths of stay [LOSs] in the acute care and rehabilitation hospital, physical therapy care, discharge destination, and follow-up care provisions) were gathered. The standardized response mean (SRM) was calculated for total, motor, and cognitive FIM scores to provide insight into the effect size and the responsiveness of the FIM for the patients with WNV in this study. RESULTS: All patients were

admitted to the rehabilitation hospital from acute care hospitals following LOSs ranging from 1 to 62 days. The rehabilitation hospital LOS ranged from 2 to 304 days. These patients had significant comorbidities including hypertension (43.75%), diabetes mellitus (41.67%), acute respiratory failure (37.5%), ventilator dependency/tracheostomy (33.33%), and pneumonia (29.17%). Their admission FIM scores ranged from 13 to 116 (mean = 45.8 +/- 28.2) and discharge FIM scores ranged from 18 to 121 (mean = 75.1 +/- 34.2). The change in FIM during inpatient rehabilitation was statistically significant (P < .001). The calculated SRM for the total (1.06) and motor (1.12) FIM indicate a large effect size, whereas the SRM for the cognitive FIM (0.79) indicates a moderate effect. The majority of patients were discharged home or to a nursing facility (46%), skilled or extended care (38%) with a need for continued rehabilitation services. DISCUSSION AND CONCLUSIONS: The manifestation of the WNV and functional outcomes after comprehensive rehabilitation vary from patient to patient. Higher numbers of comorbid conditions lead to more complex presentation and challenge rehabilitation professionals to design individualized plans of care to enable these patients to achieve the highest functional outcomes. Most patients require follow-up physical therapy care after discharge from rehabilitation

Hsieh, C. H., DeJong, G., Groah, S., Ballard, P. H., Horn, S. D., & Tian, W. (2013). Comparing rehabilitation services and outcomes between older and younger people with spinal cord injury. *Arch Phys Med Rehabil, 94*(4 Suppl), S175-186. doi:10.1016/j.apmr.2012.10.038

OBJECTIVE: To compare patient and injury characteristics, rehabilitation services, and outcomes between people incurring traumatic spinal cord injury (SCI) at younger and older ages. DESIGN: Multisite prospective observational cohort study. SETTING: Six acute rehabilitation facilities. PARTICIPANTS: Patients (N=866) aged >/= 16 years admitted to participating centers for their initial rehabilitation after SCI. INTERVENTIONS: Not applicable. MAIN OUTCOME MEASURES: Motor FIM scores at discharge and 1-year postinjury, discharge location, and post acute clinical pathways. RESULTS: Patients were divided into 4 age-at-injury groups: 16 to 29, 30 to 44, 45 to 60, and >60 years of age. Older adults (>60 y) incurring SCI were more likely to be married, retired/unemployed, on Medicare, and to have attained more education. Their injuries mostly resulted from falls and were incomplete in nature. The oldest group had the highest severity of illness, lowest admission and discharge motor FIM scores, and longer rehabilitation stay. They received relatively less rehabilitation than younger groups. They spent proportionately more time in occupational therapy working on preparatory activities and less time on self-care activities during inpatient rehabilitation. In the aged >60 years group, 80% went home at discharge; 17.2% were discharged to a nursing home. Younger groups were less likely to go to a nursing home. Admission motor FIM was the most significant predictor of motor FIM at discharge and 1-year anniversary across age groups. But the age groups differed significantly in patient and treatment factors that explained their respective outcomes. CONCLUSIONS: Older injured individuals experienced a different clinical pathway from younger patients. The present study suggests the need for development of a rehabilitation program tailored specifically to older adults.

Jung HY, Trivedi AN, Grabowski DC, Mor V. 2016. Does More Therapy in Skilled Nursing Facilities Lead to Better Outcomes in Patients with Hip Fracture? Phys Ther Jan;96(1):81-9.

Background:  Skilled nursing facilities (SNFs) have increasingly been providing more therapy hours to beneficiaries of Medicare. It is not known whether these increases have improved patient outcomes. Objective:  The study objectives were: (1) to examine temporal trends in therapy hour volumes and (2) to evaluate whether more therapy hours are associated with improved patient outcomes. Design: This was a retrospective cohort study Methods: Data sources included the Minimum Data Set, Medicare inpatient claims, and the Online Survey, Certification, and Reporting System. The study population consisted of 481,908 beneficiaries of Medicare fee-for-service who were admitted to 15,496 SNFs after hip fracture from 2000 to 2009. Linear regression models with facility and time fixed effects were used to estimate the association between the quantity of therapy provided in SNFs and the likelihood

of discharge to home. Results: The average number of therapy hours increased by 52% during the study period, with relatively little change in case mix at SNF admission. An additional hour of therapy per week was associated with a 3.1- percentage-point (95% confidence interval=3.0, 3.1) increase in the likelihood of discharge to home. The effect of additional therapy decreased as the Resource Utilization Group category increased, and additional therapy did not benefit patients in the highest Resource Utilization Group category. Limitations: Minimum Data Set assessments did not cover details of therapeutic interventions throughout the entire SNF stay and captured only a 7-day retrospective period for measures of the quantity of therapy provided. Conclusions: Increases in the quantity of therapy during the study period cannot be explained by changes in case mix at SNF admission. More therapy hours in SNFs appear to improve outcomes, except for patients with the greatest need.

Kroll, C., & Fisher, T. (2018). Justifying Rehabilitation Intensity Through Functional Performance Measures in Postacute Care. *Am J Occup Ther, 72*(1), 7201090010p7201090011-7201090010p7201090016. doi:10.5014/ajot.2018.721002

The Centers for Medicare and Medicaid Services (CMS) has scrutinized the provision of rehabilitation services in skilled nursing facilities (SNFs) for some time. Little research guidance exists on appropriate dosage or rehabilitation intensity (RI) among SNF patients or patients in other postacute care (PAC) settings. CMS developed a PAC assessment, the Continuity Assessment Record and Evaluation (CARE) Tool, in response to questions about what issues drive placement in various PAC settings under Medicare. The ability to adequately assess functional outcomes and correlate them to the RI provided by using the CARE Tool is promising. However, further research, policy advocacy, and practice analysis must be undertaken to promote and protect adequate access to occupational therapy and physical therapy in SNFs and other PAC settings. Individual practitioners must participate in data gathering to ensure that the data for analysis are fully informed by the occupational therapy perspective.

Morghen, S., Morandi, A., Guccione, A. A., Bozzini, M., Guerini, F., Gatti, R., . . . Bellelli, G. (2017). The association between patient participation and functional gain following inpatient rehabilitation. *Aging Clin Exp Res, 29*(4), 729-736. doi:10.1007/s40520-016-0625-3

OBJECTIVES: To evaluate patients' participation during physical therapy sessions as assessed with the Pittsburgh rehabilitation participation scale (PRPS) as a possible predictor of functional gain after rehabilitation training. METHODS: All patients aged 65 years or older consecutively admitted to a Department of Rehabilitation and Aged Care (DRAC) were evaluated on admission regarding their health, nutritional, functional and cognitive status. Functional status was assessed with the functional independence measure (FIM) on admission and at discharge. Participation during rehabilitation sessions was measured with the PRPS. Functional gain was evaluated using the Montebello rehabilitation factor score (MRFS efficacy), and patients stratified in two groups according to their level of functional gain and their sociodemographic, clinical and functional characteristics were compared. Predictors of poor functional gain were evaluated using a multivariable logistic regression model adjusted for confounding factors. RESULT: A total of 556 subjects were included in this study. Patients with poor functional gain at discharge demonstrated lower participation during physical therapy sessions were significantly older, more cognitively and functionally impaired on admission, more depressed, more comorbid, and more frequently admitted for cardiac disease or immobility syndrome than their counterparts. There was a significant linear association between PRPS scores and MRFS efficacy. In a multivariable logistic regression model, participation was independently associated with functional gain at discharge (odds ratio 1.51, 95 % confidence interval 1.19-1.91). CONCLUSION: This study showed that participation during physical therapy affects the extent of functional gain at discharge in a large population of older patients with multiple diseases receiving in-hospital rehabilitation.

Nehra, D., Nixon, Z. A., Lengenfelder, C., Bulger, E. M., Cuschieri, J., Maier, R. V., & Arbabi, S. (2016). Acute Rehabilitation after Trauma: Does it Really Matter? *J Am Coll Surg, 223*(6), 755-763. doi:10.1016/j.jamcollsurg.2016.09.001

BACKGROUND: The impact of post-discharge rehabilitation care for the trauma patient remains poorly investigated. Here we describe the functional outcomes of trauma patients discharged to an inpatient rehabilitation facility (IRF), and compare the likelihood of discharge home, 1-year rehospitalization, and 1-year mortality between patients discharged to an IRF and a propensity score-matched cohort of patients not discharged to an IRF. STUDY DESIGN: The Washington State Rehabilitation Registry was used to collect data for all trauma patients discharged to an IRF between 2011 and 2012. These charts were linked to the Washington State Trauma Registry and the Comprehensive Hospital Abstract Reporting System database to obtain detailed patient, injury, and mortality data. Propensity score matching was used to identify a control group of patients who were not discharged to an IRF. Primary outcomes measures were improvement in Functional Independence Measure score with inpatient rehabilitation and the likelihood of discharge home, 1-year rehospitalization, and 1-year mortality. RESULTS: Nine hundred and thirty-three trauma patients were discharged to an IRF between 2011 and 2012. Total functional independence measure scores improved from 63.7 (SD 20.3) to 92.2 (SD 20.9) ($p < 0.001$) with care at an IRF. When patients discharged to an IRF were compared with the propensity score-matched control patients, rehabilitation was found to significantly increase the likelihood of discharge to home (odds ratio = 9.41; 95% CI, 6.80-13.01) and to decrease 1-year mortality (odds ratio = 0.60; 95% CI, 0.39-0.92). CONCLUSIONS: Acute trauma patients should be recognized as an underserved population that would benefit considerably from inpatient rehabilitation services after discharge from the hospital.

O'Brien, S. R., & Xue, Y. (2016). Inpatient Rehabilitation Outcomes in Patients With Stroke Aged 85 Years or Older. *Phys Ther, 96*(9), 1381-1388. doi:10.2522/ptj.20150364

BACKGROUND: In the United States, people 85 years of age or older have a growing number of strokes each year, and this age group is most at risk for disability. Inpatient rehabilitation facilities (IRFs) adhere closest to post-acute stroke rehabilitation guidelines and have the most desirable outcomes compared with skilled nursing facilities. As stroke is one of the leading causes of disability, knowledge of postrehabilitation outcomes is needed for this age group, although at present such information is limited. OBJECTIVE: The purpose of this study was to describe functional and discharge outcomes after IRF rehabilitation in people with stroke aged 85 years or older. DESIGN: A serial, cross-sectional design was used. METHODS: Inpatient Rehabilitation Facility-Patient Assessment Instrument data were analyzed beginning in 2002 for the first 5.5 years after implementation of the prospective payment system and included 71,652 cases. Discharge function, measured using the Functional Independence Measure (FIM), and community discharge were the discharge outcome measures. Sample description used frequencies and means. Generalized estimating equations (GEEs) with post hoc testing were used to analyze the annual trends for discharge FIM and community discharge by age group (85-89, 90-94, 95-99, and >/=100 years). Risk-adjusted linear and logistic GEE models, with control for cluster, were used to analyze the association between both outcome measures and age group. RESULTS: Over 5.5 years, mean discharge FIM scores decreased by 3.6 points, and mean achievement of community discharge decreased 5.5%. Approximately 54% of the sample achieved community discharge. Continuous and logistic GEEs revealed factors associated with discharge outcomes. LIMITATIONS: Results obtained using an observational design should not be viewed as indicating causation. The lack of control for a caregiver may have altered results. CONCLUSIONS: The very elderly people admitted to IRF stroke rehabilitation made functional gains, and most were able to return to the community.

Rosenbaum, A. M., Gordon, W. A., Joannou, A., & Berman, B. A. (2018). Functional outcomes following post-acute rehabilitation for moderate-to-severe traumatic brain injury. *Brain Inj, 32*(7), 907-914. doi:10.1080/02699052.2018.1469040

OBJECTIVE: The objective of this study was to examine the benefits of long-term inpatient rehabilitation for individuals with moderate-to-severe traumatic brain injuries (TBIs). METHODS: Retrospective database review of 67 individuals with moderate-to-severe TBI admitted to a specialised inpatient TBI program. Outcome measures are as follows: (1) functional independence measure +

functional assessment measure (FIM+FAM; admission, discharge, change scores); (2) discharge designation (community vs. long-term care (LTC)). RESULTS: There was a mean improvement on FIM+FAM of 54.19 points (SD = 35.63) or 67% between admission and discharge (t(66) = -12.45, p < 0.001). Mean time post-injury upon completion of therapy was 409.59 days (SD = 343.93). Upon completion of rehabilitation, 50 (75%) participants were discharged to community and 17 to LTC. Among those returning to community, those with longer length of stays were more severely disabled on admission (t(35.9) = -4.86, p < 0.001). Controlling for admission functional status, individuals returning to community following >90 days of therapy required a mean of 378.94 days (SD = 298.86) to achieve comparable gains to those less impaired who received shorter periods of rehabilitation (F(1) = 0.530, p = 0.47). CONCLUSION: Continued specialised inpatient services following acute inpatient rehabilitation for individuals with moderate-to-severe TBI can reduce the level of dependency and enhance the likelihood of return to community living.

Sauter et al (2013). Functional outcomes of persons who underwent dysvascular lower extremity amputations: effect of postacute rehabilitation setting. Am J Phys Med Rehabil. Apr;92(4):287-96.

OBJECTIVE: The aim of this study was to examine the effect of postacute rehabilitation setting on functional outcomes among patients who underwent major dysvascular lower extremity amputations. DESIGN: This is a population-based prospective cohort study conducted in Maryland and Wisconsin. Data collected from medical records and patient interviews conducted during acute hospitalization after amputation and at 6 mos after the acute care discharge were analyzed using multivariate models and instrumental variable techniques. RESULTS: A total of 297 patients were analyzed on the basis of postacute care rehabilitation setting: acute inpatient rehabilitation facility (IRF), skilled nursing facility (SNF), or home. The majority (43.4%) received care in an IRF; 32%, in an SNF; and 24.6%, at home. On the Short Form-36 subscales, significantly improved outcomes were observed for the patients receiving postacute care at an IRF relative to those cared for at an SNF in physical function, role physical, and physical component summary score. Patients receiving postacute care in IRFs also experienced better role physical and physical component summary score outcomes compared with those discharged directly home. In addition, patients receiving postacute care in an IRF were significantly more likely to score in the top quartile for general health in IRF compared with SNF or home and less likely to score in the lowest quartile for physical function, role physical, and physical component summary score in IRF compared with SNF. Lower activity of daily living impairment was observed in IRF compared with SNF. CONCLUSIONS: Among this large and diverse cohort of patients who underwent major dysvascular lower limb amputations, receipt of interdisciplinary rehabilitation services in an IRF yielded improved functional outcomes 6 mos after amputation relative to care received in SNFs or at home.

Siebens, H. C., Sharkey, P., Aronow, H. U., Deutscher, D., Roberts, P., Munin, M. C., . . . Horn, S. D. (2016). Variation in Rehabilitation Treatment Patterns for Hip Fracture Treated With Arthroplasty. PM&R, *8*(3), 191-207.

BACKGROUND: Recommendations for health care redesign often advocate for comparative effectiveness research that is patient-centered. For patients who require rehabilitation services, a first step in this research process is to understand current practices for specific patient groups. OBJECTIVE: To document in detail the physical and occupational therapy treatment activities for inpatient hip fracture rehabilitation among 3 patient subgroups distinguished by their early rate of functional recovery between time of surgery to rehabilitation admission. DESIGN: Multicenter prospective observational cohort, practice-based evidence, study. SETTING: Seven skilled nursing facilities and 11 inpatient rehabilitation facilities across the United States. PARTICIPANTS: A total of 226 patients with hip fractures treated with hip arthroplasty. METHODS: Comparisons of physical and occupational therapy treatment activities among 3 groups with different initial recovery trajectory (IRT) rates (slower, moderate, faster). MAIN OUTCOME MEASURE(S): Percent of patients in each IRT group exposed to each physical and occupational therapy activity (exposure), and mean minutes per week

for each activity (intensity). RESULTS: The number of patients exposed to different physical or occupational therapy activities varied within the entire sample. More specifically, among the 3 IRT groups, significant differences in exposure occurred for 44% of physical therapy activities and 39% of occupational therapy activities. More patients in the slower recovery group, IRT 1, received basic activities of daily living treatments and more patients in the faster recovery group, IRT 3, received advanced activities. The moderate recovery group, IRT 2, had some treatments similar to IRT 1 group and others similar to IRT 3 group. CONCLUSIONS: Analyses of practice-based evidence on inpatient rehabilitation of hip fracture patients treated with arthroplasty identified differences in therapy activities among three patient groups classified by IRT rates. These results may enhance physiatrists', other physicians', and rehabilitation teams' understanding of inpatient rehabilitation for these patients and help design future comparative effectiveness research.

Skidmore, E. R., Dawson, D. R., Butters, M. A., Grattan, E. S., Juengst, S. B., Whyte, E. M., . . . Becker, J. T. (2015). Strategy Training Shows Promise for Addressing Disability in the First 6 Months After Stroke. *Neurorehabil Neural Repair, 29*(7), 668-676. doi:10.1177/1545968314562113

OBJECTIVE: To examine the feasibility of a strategy training clinical trial in a small group of adults with stroke-related cognitive impairments in inpatient rehabilitation, and to explore the impact of strategy training on disability. DESIGN: Non-randomized two-group intervention pilot study. SETTING: Two inpatient rehabilitation units within an academic health centre. PARTICIPANTS: Individuals with a primary diagnosis of acute stroke, who were admitted to inpatient rehabilitation and demonstrated cognitive impairments were included. Individuals with severe aphasia; dementia; major depressive disorder, bipolar, or psychotic disorder; recent drug or alcohol abuse; and anticipated length of stay less than five days were excluded. INTERVENTION: Participants received strategy training or an attention control session in addition to usual rehabilitation care. Sessions in both groups were 30-40 minutes daily, five days per week, for the duration of inpatient rehabilitation. MAIN OUTCOME MEASURES: We assessed feasibility through participants' recruitment and retention; research intervention session number and duration; participants' comprehension and engagement; intervention fidelity; and participants' satisfaction. We assessed disability at study admission, inpatient rehabilitation discharge, 3 and 6 months using the Functional Independence Measure. RESULTS: Participants in both groups (5 per group) received the assigned intervention (>92% planned sessions; >94% fidelity) and completed follow-up testing. Strategy training participants in this small sample demonstrated significantly less disability at six months (M (SE) = 117 (3)) than attention control participants (M(SE) = 96 (14); t 8 = 7.87, P = 0.02). CONCLUSIONS: It is feasible and acceptable to administer both intervention protocols as an adjunct to acute inpatient rehabilitation, and strategy training shows promise for reducing disability.

Wang, H., Camicia, M., Terdiman, J., Mannava, M. K., Sidney, S., & Sandel, M. E. (2013). Daily treatment time and functional gains of stroke patients during inpatient rehabilitation. PM&R *5*(2), 122-128. doi:10.1016/j.pmrj.2012.08.013

OBJECTIVE: To study the effects of daily treatment time on functional gain of patients who have had a stroke. DESIGN: A retrospective cohort study. SETTING: An inpatient rehabilitation hospital (IRH) in northern California. PARTICIPANTS: Three hundred sixty patients who had a stroke and were discharged from the IRH in 2007. INTERVENTIONS: Average minutes of rehabilitation therapy per day, including physical therapy, occupation therapy, speech and language therapy, and total treatment. MAIN OUTCOME MEASURES: Functional gain measured by the Functional Independence Measure, including activities of daily living, mobility, cognition, and the total of the Functional Independence Measure (FIM) scores. RESULTS: The study sample had a mean age of 64.8 years; 57.4% were men and 61.4% were white. The mean total daily therapy time was 190.3 minutes, and the mean total functional gain was 26.0. A longer daily therapeutic duration was significantly associated with total functional gain (r = .23, P = .0094). Patients who received a total therapy time of <3.0 hours per day had significantly lower total functional gain than did those treated >/=3.0 hours. No significant

difference in total functional gain was found between patients treated >/=3.0 but <3.5 hours and >/=3.5 hours per day. The daily treatment time of physical therapy, occupational therapy, and speech and language therapy also was significantly associated with corresponding subscale functional gains. In addition, hemorrhagic stroke, left brain injury, earlier IRH admission, and a longer IRH stay were associated with total functional improvement. CONCLUSIONS: The study demonstrated a significant relationship between daily therapeutic duration and functional gain during IRH stay and showed treatment time thresholds for optimal functional outcomes for patients in inpatient rehabilitation who had a stroke.

Winstein, C. J., Stein, J., Arena, R., Bates, B., Cherney, L. R., Cramer, S. C., . . . Zorowitz, R. D. (2016). Guidelines for Adult Stroke Rehabilitation and Recovery: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. *Stroke, 47*(6), e98-e169. doi:10.1161/str.0000000000000098

PURPOSE: The aim of this guideline is to provide a synopsis of best clinical practices in the rehabilitative care of adults recovering from stroke. METHODS: Writing group members were nominated by the committee chair on the basis of their previous work in relevant topic areas and were approved by the American Heart Association (AHA) Stroke Council's Scientific Statement Oversight Committee and the AHA's Manuscript Oversight Committee. The panel reviewed relevant articles on adults using computerized searches of the medical literature through 2014. The evidence is organized within the context of the AHA framework and is classified according to the joint AHA/American College of Cardiology and supplementary AHA methods of classifying the level of certainty and the class and level of evidence. The document underwent extensive AHA internal and external peer review, Stroke Council Leadership review, and Scientific Statements Oversight Committee review before consideration and approval by the AHA Science Advisory and Coordinating Committee. RESULTS: Stroke rehabilitation requires a sustained and coordinated effort from a large team, including the patient and his or her goals, family and friends, other caregivers (eg, personal care attendants), physicians, nurses, physical and occupational therapists, speech-language pathologists, recreation therapists, psychologists, nutritionists, social workers, and others. Communication and coordination among these team members are paramount in maximizing the effectiveness and efficiency of rehabilitation and underlie this entire guideline. Without communication and coordination, isolated efforts to rehabilitate the stroke survivor are unlikely to achieve their full potential. CONCLUSIONS: As systems of care evolve in response to healthcare reform efforts, postacute care and rehabilitation are often considered a costly area of care to be trimmed but without recognition of their clinical impact and ability to reduce the risk of downstream medical morbidity resulting from immobility, depression, loss of autonomy, and reduced functional independence. The provision of comprehensive rehabilitation programs with adequate resources, dose, and duration is an essential aspect of stroke care and should be a priority in these redesign efforts.

Wysocki, A., Thomas, K. S., & Mor, V. (2015). Functional Improvement Among Short-Stay Nursing Home Residents in the MDS 3.0. *J Am Med Dir Assoc, 16*(6), 470-474. doi:10.1016/j.jamda.2014.11.018

OBJECTIVES: To examine the completeness of the activities of daily living (ADL) items on admission and discharge assessments and the improvement in ADL performance among short-stay residents in the newly adopted Minimum Data Set (MDS) 3.0. DESIGN: Retrospective analysis of MDS admission and discharge assessments. SETTING: Nursing homes from July 1, 2011, to June 30, 2012. PARTICIPANTS: New nursing home residents admitted from acute hospitals with corresponding admission and discharge assessments between July 1, 2011, and June 30, 2012, who had a length of stay of 100 days or less. MEASUREMENTS: ADL self-performance items, including bed mobility, transfer, walking in room, walking in corridor, locomotion on unit, locomotion off unit, dressing, eating, toilet use, and personal hygiene, at admission and discharge. RESULTS: The ADL self-performance items are complete at both admission and discharge, with less than 1% missing for any item. More than 60% of residents improved over the course of their post-acute stay. New short-stay nursing home residents with conditions such as cognitive impairment, delirium, dementia, heart failure, and stroke showed less

improvement in ADL performance during their stay. CONCLUSION: The discharge assessment data in the MDS 3.0 provide new information to researchers and providers to examine and track ADL performance. Nursing homes can identify and track patients who require more intensive therapies or targeted interventions to achieve functional improvement during their stay. Future research can examine facility-level measures to better understand how ADL improvement varies across facilities.

**Scoping Review Methodology**

To prepare for the NQF Endorsement Maintenance Review for this measure, we sought to identify relevant literature since our 2014 NQF submission. The literature search focused on how one intervention/service, therapy, is associated with the measure, functional outcomes. Therapy is one of the processes listed in the Structure-Process-Outcome Model (see Figure 1a). This model shows that IRF staff, including therapists, can implement interventions that result in improving their patients' functional outcomes, specifically their mobility and self-care outcomes.

Our team conducted a scoping review that included a systematic search of published literature relevant to our IRF measures (NQF #2633, NQF #2634, NQF #2635, and NQF #2636). To identify the relevant literature, we identified the search strategy with input from all team members. The search strategy included relevant terms for the setting, interventions and outcomes that align with these IRF measures. We included articles that met all three criteria. Below, we outline our search strategy. Note these are only examples and are not fully comprehensive of the search terminology we used.

1. **Setting search terms:** IRFs are the primary setting of focus as these measures assess patient functional outcomes (mobility and self-care) in IRFs. We used a variety of terms that are commonly used to describe IRFs such as, "inpatient rehabilitation facility" "rehabilitation centers" or "intensive rehabilitation". We also included searches for articles about "Skilled Nursing Facilities" or "SNFs" or "short-stay nursing home" as SNFs offer similar rehabilitation treatments as IRFs. More generally, we also searched for "post-acute care settings" as some research articles focus on post-acute care (PAC) settings and may be relevant to IRFs or SNFs.
2. **Intervention search terms:** We searched a variety of key terms such as "therapy" or "mobilization" or "intervention".
3. **Functional outcomes:** We used key words such as: "functional outcome" or "functional improvement" or "activities of daily living".

Exclusion criteria were pre-determined by the team before the search was conducted. Exclusion criteria were: any articles published before January 1, 2013; any articles published outside of the US that did not use US based data; articles not written in English; articles not focused on human outcomes; and any articles that were focused on Long-term Care Hospitals or LTCHs, and other publication types that were not research-based such as opinion pieces or commentaries were excluded. We also included 6 additional articles we found that meet our inclusion criteria that were not identified by our PubMed search.

Our initial search yielded 181 articles. For every publication identified, we assigned two coders to independently review each abstract to determine if the study was relevant (i.e., should be included or excluded). The team met to compare decisions, and for abstracts for which we disagreed, the team re-reviewed the abstract together and we made a consensus decision. All Case Reports were excluded, because these articles focus on one individual and the findings are not generalizable. We also excluded articles for other reasons including those that describe outcomes that are not a focus of our measures, including cognition outcomes, readmissions or discharge destination. In addition, we excluded those articles focused on outpatient or acute care settings.

Our final scoping review results yielded 26 articles for inclusion. Following our inclusion decisions, we also grouped the articles by type of setting (IRF, SNF, IRF and SNF, or other), of functional outcome (self-care, mobility, motor function), and if the study focused on a specific diagnosis (e.g. stroke) or multiple diagnoses. Twenty-two articles were included that included interventions or outcomes for motor function (mobility, self-care, and bladder); no articles were identified that focused on self-care outcomes.

**Previous NQF Submission 2014**

Treatments furnished by IRF clinicians focus on reducing patients' impairments and activity limitations as well as managing patients' medical, psychological and other health needs. The relationship between rehabilitation interventions and patients' functional outcomes has been challenging to examine (Foley et al., 2012), because rehabilitation interventions tend to be multidisciplinary, tailored to each patient's specific needs and there are no standardized definitions and no standardized measurement of interventions. In addition, research examining the optimal "dose" of therapy has been limited in IRFs due, in part, to the provision of intensive therapy services to all patients, and concern about the lack of variability in the amount of therapy provided. The rehabilitation treatment-outcome knowledge gap is recognized, and several efforts are underway to classify interventions using standardized terminology in order to better understand the relationship between interventions and outcomes; that is, the active ingredients of a rehabilitation program (Natale et al., 2009; Ozelie et al., 2009; Johnson et al., 2009; Rundquist et al., 2011; Taylor-Schroeder, 2011). Several studies have examined the therapy dose-outcome relationship, and reported higher amounts of therapy were associated with better functional improvement (Jette, Warren & Wirtalla, 2010; Lenze et al., 2012; Ozelie et al., 2012; Wang et al., 2013; Mallinson et al, 2014; Lohse, Lang & Boyd, 2014). In addition, O'Brien, Xue, Ingersoll & Kelly (2013) reported that shorter IRF stays were associated with lower patient functioning at discharge; the average IRF length of stays decreased 1.8 days between 2002 and 2007, and the patients in 2007 had lower functional abilities at discharge compared to patients in 2002.

**Citations**


Foley, N., Pereira, S., Salter, K., Meyer, M., McClure, J. A., & Teasell, R., (2012). Are recommendations regarding inpatient therapy intensity following acute stroke really evidence-based? *Topics in Stroke Rehabilitation*. 19(2):96-103.

Jette, D. U., R. L. Warren, & C. Wirtalla. (2005). The relation between therapy intensity and outcomes of rehabilitation in skilled nursing facilities. *Archives of Physical Medicine and Rehabilitation,* 86 (3), 373-9.

Johnson K., Bailey J., Rundquist J., Dimond P., McDonald CA., Reyes IA., … Gassaway J. (2009). SCIRehab Project series: the supplemental nursing taxonomy. *Journal of Spinal Cord Medicine*. 32(3):329-35.

Lenze, E. J., Host, H. H., Hildebrand M. W., Morrow-Howell, N., Carpenter, B., Freedland, K. E., … Binder, E, F. (2012). Enhanced medical rehabilitation increases therapy intensity and engagement and improves functional outcomes in postacute rehabilitation of older adults: a randomized-controlled trial. *Journal of the American Medical Directors Association*. 13(8):708-12.

Lohse, K. R., Lang, C. E., & Boyd, L. A. (2014). Is more better? Using metadata to explore dose-response relationships in stroke rehabilitation. *Stroke*. 45(7):2053-8.

Mallinson, T., Deutsch, A., Bateman, J., Tseng, H. Y., Manheim, L., Almagor, O., Heinemann, A., W. (2014). Comparison of discharge functional status after rehabilitation in skilled nursing, home health, and medical rehabilitation settings for patients after hip fracture repair. *Archives of Physical Medicine & Rehabilitation*. 95(2):209-17.

Natale A., Taylor S., LaBarbera J., Bensimon L., McDowell S., Mumma S.L., … Gassaway J. (2009). SCIRehab Project series: the physical therapy taxonomy. *Journal of Spinal Cord Medicine*. 32(3):270-82.

O'Brien, S. R., Xue, Y., Ingersoll, G., & Kelly, A. (2013). Shorter length of stay is associated with worse functional outcomes for Medicare beneficiaries with stroke. *Physical Therapy, 93,* 1592–1602.

Ozelie R., Sipple C., Foy T., Cantoni K., Kellogg K., Lookingbill J., … Gassaway J. (2009).

SCIRehab Project series: the occupational therapy taxonomy. *Journal of Spinal Cord Medicine*. 32(3):283-97.

Ozelie R., Gassaway J., Buchman E., Thimmaiah D., Heisler L., Cantoni K., … Whiteneck G. (2012). Relationship of occupational therapy inpatient rehabilitation interventions and patient characteristics to outcomes following spinal cord injury: the SCIRehab project. *Journal of Spinal Cord Medicine*. 35(6):527-46.

Rundquist J., Gassaway J., Bailey J., Lingefelt P., Reyes IA., & Thomas J. (2011). The SCIRehab project: treatment time spent in SCI rehabilitation. Nursing bedside education and care management time during inpatient spinal cord injury rehabilitation. *Journal of Spinal Cord Medicine*. 34(2):205-15.

Taylor-Schroeder S., LaBarbera J., McDowell S., Zanca J.M., Natale A., Mumma S., … Backus D. (2011). The SCIRehab project: treatment time spent in SCI rehabilitation. Physical therapy treatment time during inpatient spinal cord injury rehabilitation. *Journal of Spinal Cord Medicine*. 34(2):149-61.

Wang, H., Camicia, M., Terdiman, J., Mannava, M. K., Sidney, S., & Sandel, M. E. (2013). Daily treatment time and functional gains of stroke patients during inpatient rehabilitation. *Physical Medicine & Rehabilitation, 5*(2), 122-128.

We identified evidence from literature searches using PubMed and the Cumulative Index of Nursing and Allied Health Literature (CINAHL) and in reviews of references cited in the relevant identified studies.

1. Foley, N., Pereira, S., Salter, K., Meyer, M., McClure, J. A., & Teasell, R., (2012). Are recommendations regarding inpatient therapy intensity following acute stroke really evidence-based? *Topics in Stroke Rehabilitation*. 19(2):96-103.

   Six clinical practice guidelines were retrieved and examined to determine what recommendation, if any, had been made regarding the daily provision of therapy during inpatient rehabilitation. All studies cited by the guideline authors to support their recommendations were identified and retrieved. Studies in which treatment was (a) focused on motor recovery, (b) initiated during inpatient rehabilitation, and (c) provided within 3 months of stroke onset were reviewed in greater detail. Three of the 6 identified guidelines recommended daily minimum amounts of therapy, ranging from 45 to 60 minutes each day of occupational and physiotherapy, and 3 made general statements indicating that increased intensity of therapy was either recommended or was not recommended. We believe the evidence base cannot support a specific recommendation related to therapy intensity during inpatient rehabilitation following stroke.

2. Jette, D. U., R. L. Warren, & C. Wirtalla. (2005). The relation between therapy intensity and outcomes of rehabilitation in skilled nursing facilities. *Archives of Physical Medicine and Rehabilitation,* 86 (3), 373-9.

   The aim of the study is to examine the relation between therapy intensity, including physical therapy, occupational therapy, and speech and language therapy, provided in a skilled nursing facility setting and patients' outcomes. Higher physical therapy and occupational therapy intensities were associated with greater odds of improving by at least 1 stage in mobility and activities of daily living functional independence across each condition. The speech and language therapy intensity was associated with improved motor and executive control functional stages for patients with stroke. Therapy intensities accounted for small proportions of model variances in all outcomes. Higher therapy intensity was associated with better outcomes as they relate to LOS and functional improvement for patients who have stroke, orthopedic conditions, and cardiovascular and pulmonary conditions and are receiving rehabilitation in skilled nursing facilities.

3. Johnson K., Bailey J., Rundquist J., Dimond P., McDonald CA., Reyes IA., … Gassaway J. (2009). SCIRehab Project series: the supplemental nursing taxonomy. *Journal of Spinal Cord Medicine*. 32(3):329-35.
   Spinal cord injury rehabilitation nurses document the occurrence of educational and care management efforts in traditional nursing documentation methods but not the intensity (or dose) of such interactions. This article describes a process to capture these nursing interventions. Nurses at 6 US inpatient spinal cord injury centers developed a taxonomy of nursing patient education efforts and care management. This was subsequently incorporated into a point-of-care documentation system and used to capture details of nursing care for 1,500 Spinal cord injury rehabilitation patients. The taxonomy consists of 10 education and 3 care management categories. The point-of-care system includes time spent on each category along with an indication of whether the patient and/or family received the education/care management. In addition, a subjective measure of patient participation in nursing activities is included.
4. Lenze, E. J., Host, H. H., Hildebrand M. W., Morrow-Howell, N., Carpenter, B., Freedland, K. E., Binder, E, F. (2012). Enhanced medical rehabilitation increases therapy intensity and engagement and improves

functional outcomes in postacute rehabilitation of older adults: a randomized-controlled trial. *Journal of the American Medical Directors Association*. 13(8):708-12.

This study tested Enhanced Medical Rehabilitation, an intervention designed to increase patient engagement in, and intensity of, daily physical and occupational therapy sessions in a skilled nursing facility. This was a randomized controlled trial of Enhanced Medical Rehabilitation versus standard-of-care rehabilitation. Participants were 26 older adults admitted from a hospital for postacute rehabilitation. Participants randomized to Enhanced Medical Rehabilitation had higher intensity therapy and were more engaged in their rehabilitation sessions; they had more improvement in gait speed and 6-minute walk, with a trend for better improvement of Barthel Index, compared with participants randomized to standard-of-care rehabilitation. Higher intensity and patient engagement in the postacute rehabilitation setting is achievable, with resultant better functional outcomes for older adults.

5. Lohse, K. R., Lang, C. E., & Boyd, L. A. (2014). Is more better? Using metadata to explore dose-response relationships in stroke rehabilitation. *Stroke*. 45(7):2053-8.

   The primary objective of this meta-analysis was to explore the relationship between time scheduled for therapy and improvement in motor therapy for adults after stroke by (1) comparing high doses to low doses and (2) using metaregression to quantify the dose-response relationship further. Databases were searched to find randomized controlled trials that were not dosage matched for total time scheduled for therapy. Regression models were used to predict improvement during therapy as a function of total time scheduled for therapy and years after stroke. Overall, treatment groups receiving more therapy improved beyond control groups that received less. There is a positive relationship between the time scheduled for therapy and therapy outcomes.

6. Mallinson, T., Deutsch, A., Bateman, J., Tseng, H. Y., Manheim, L., Almagor, O., Heinemann, A., W. (2014). Comparison of discharge functional status after rehabilitation in skilled nursing, home health, and medical rehabilitation settings for patients after hip fracture repair. *Archives of Physical Medicine & Rehabilitation*. 95(2):209-17.

   The aim of this study was to examine differences in rehabilitation outcomes across 3 post-acute care rehabilitation settings for patients after hip fracture repair. Participants were patients (N=181) receiving rehabilitation following hip fracture. Inpatient rehabilitation facility and home health agency patients had lower self-care function at discharge relative to skilled nursing facility patients controlling for patient characteristics, severity, comorbidities, and services. Inpatient rehabilitation facility and skilled nursing facility patients received about the same total minutes of therapy over their PAC stays (~2100 min on average), whereas home health patients received only approximately 25% as many minutes. Setting-specific effects varied depending on whether self-care or mobility was the outcome of focus.

7. Natale A., Taylor S., LaBarbera J., Bensimon L., McDowell S., Mumma S.L., … Gassaway J. (2009). SCIRehab Project series: the physical therapy taxonomy. *Journal of Spinal Cord Medicine*. 32(3):270-82.

   The objective of this study was to describe a taxonomy (system to categorize and classify interventions) to examine the effects of physical therapy interventions on rehabilitation outcomes. Physical therapy clinicians and researchers from 6 centers developed a taxonomy to describe details of each PT session. The physical therapy taxonomy consists of 19 treatment activities (e. g., bed mobility, transfers, wheelchair mobility, strengthening and stretching exercises) and supplementary information to describe the associated therapeutic interventions. The detailed physical therapy taxonomy documentation process, which offers efficiency in data collection, is being used for all physical therapy sessions with 1,500 patients with acute traumatic spinal cord injury at the 6 participating centers.

8. O'Brien, S. R., Xue, Y., Ingersoll, G., & Kelly, A. (2013). Shorter length of stay is associated with worse functional outcomes for Medicare beneficiaries with stroke. *Physical Therapy, 93,* 1592–1602.

   This study examined the trends and associations between length of stay and discharge outcomes in Medicare beneficiaries with stroke treated in IRFs. Medicare beneficiaries with stroke treated in IRFs

experienced shorter length of stay, had worsening admission and discharge function, and had fewer community discharges. Worsening admission function and shorter length of stay may contribute to worsening discharge outcomes, which may indicate a lack of readiness for IRF treatment and that facility-level factors may be playing a role in shorter length of stay.

9. Ozelie R., Sipple C., Foy T., Cantoni K., Kellogg K., Lookingbill J., … Gassaway J. (2009). SCIRehab Project series: the occupational therapy taxonomy. *Journal of Spinal Cord Medicine*. 32(3):283-97.

   Occupational therapy clinicians and researchers from 6 spinal cord injury rehabilitation centers developed a taxonomy to describe details of each occupational therapy session. The occupational therapy taxonomy consists of 26 occupational therapy activities (e. g., training on activities of daily living, communication, home management skills, wheelchair mobility, bed mobility, transfers, balance, strengthening, stretching, equipment evaluation, and community reintegration). Treatment descriptions are enhanced further with identification of assistance needs, patient direction of care, and family involvement, which help to describe and guide occupational therapy activity selection. The electronic documentation system is being used at 6 centers for all occupational therapy sessions with 1,500 patients with acute traumatic spinal cord injury.

10. Ozelie R., Gassaway J., Buchman E., Thimmaiah D., Heisler L., Cantoni K., … Whiteneck G. (2012). Relationship of occupational therapy inpatient rehabilitation interventions and patient characteristics to outcomes following spinal cord injury: the SCIRehab project. *Journal of Spinal Cord Medicine*. 35(6):527-46.

    Occupational therapists at 6 inpatient rehabilitation centers documented detailed information about treatment provided. Occupational therapy treatment variables explain a small amount of variation in FIM outcomes for the full sample and significantly more in two functionally homogeneous subgroups. For patients with motor complete paraplegia, more time spent in clothing management and hygiene related to toileting was a strong predictor of higher scores on the lower body items of the self-care function. Among patients with motor complete low tetraplegia, higher scores for the FIM lower body self-care items were associated with more time spent on lower body dressing, manual wheelchair mobility training, and bathing training. The impact of occupational therapy treatment on functional outcomes is more evident when examining more homogeneous patient groupings and outcomes specific to the groupings.

11. Rundquist J., Gassaway J., Bailey J., Lingefelt P., Reyes IA., & Thomas J. The SCIRehab project: treatment time spent in SCI rehabilitation. Nursing bedside education and care management time during inpatient spinal cord injury rehabilitation. *Journal of Spinal Cord Medicine*. 34(2):205-15, 2011.

    Nurses providing usual care to patients with spinal cord injury documented the content and amount of time spent on each bedside interaction including details of education or care management for 42 048 shifts of nursing care. The mean number of minutes per week was 264.3. The time that nurses spent on each activity was significantly different in each neurological injury group. Fifty percent of care management time was devoted to psychosocial support, while medication, skin care, bladder, bowel, and pain management were the main education topics. Nurses in spinal cord injury rehabilitation spend a significant amount of time providing education and psychosocial support to patients and their families. Quantification of these interventions will allow researchers to discern whether there are pertinent associations between the time spent on bedside activities and patient outcomes.

12. Taylor-Schroeder S., LaBarbera J., McDowell S., Zanca J.M., Natale A., Mumma S., … Backus D. (2011). The SCIRehab project: treatment time spent in SCI rehabilitation. Physical therapy treatment time during inpatient spinal cord injury rehabilitation. *Journal of Spinal Cord Medicine*. 34(2):149-61.

    Physical therapists documented details, including time spent, of treatment provided during 37,306 physical therapy sessions that occurred during inpatient SCI rehabilitation. SCIRehab patients received a mean total of 55.3 hours of physical therapy over the course of their rehabilitation stay. Significant differences among four neurologic groups were seen in the amount of time spent on most activities, including the most common physical therapy activities of strengthening exercises, stretching, transfer training, wheelchair mobility training, and gait training. Most physical therapy work (77%) was provided in individual therapy sessions; the remaining 23% was done in group settings. Patient and injury characteristics explained only

some of the variations seen in time spent on wheelchair mobility, transfer and bed mobility training, and range of motion/stretching. Significant variation was seen in time spent on physical therapy activities within and among injury groups.

13. Wang, H., Camicia, M., Terdiman, J., Mannava, M. K., Sidney, S., & Sandel, M. E. (2013). Daily treatment time and functional gains of stroke patients during inpatient rehabilitation. *Physical Medicine & Rehabilitation, 5*(2), 122-128.

    The average total minutes of rehabilitation therapy per day, including physical therapy, occupation therapy, speech and language therapy for 360 patients who had a stroke and were discharged from the IRH in 2007 was 190.3 minutes. The mean total functional gain was 26.0. A longer daily therapeutic duration was significantly associated with total functional gain. Patients who received a total therapy time of <3.0 hours per day had significantly lower total functional gain than did those treated ≥3.0 hours. No significant difference in total functional gain was found between patients treated ≥3.0 but <3.5 hours and ≥3.5 hours per day. The daily treatment time of physical therapy, occupational therapy, and speech and language therapy also was significantly associated with corresponding subscale functional gains. The study demonstrated a significant relationship between daily therapeutic duration and functional gain during the inpatient rehabilitation facility stay and showed treatment time thresholds for optimal functional outcomes for patients in inpatient rehabilitation who had a stroke.

**1a.3. SYSTEMATIC REVIEW(S) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES, INCLUDING THOSE THAT ARE INSTRUMENT-BASED) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.**

**What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)**

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐ Other

Not Applicable. This measure is an outcome measure.

| | |
|---|---|
| **Source of Systematic Review:**<br>• **Title**<br>• **Author**<br>• **Date**<br>• **Citation, including page number**<br>• **URL** | |
| Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR. | |

| | |
|---|---|
| Grade assigned to the **evidence** associated with the recommendation with the definition of the grade | |
| Provide all other grades and definitions from the evidence grading system | |
| Grade assigned to the **recommendation** with definition of the grade | |
| Provide all other grades and definitions from the recommendation grading system | |
| Body of evidence:<br>• Quantity – how many studies?<br>• Quality – what type of studies? | |
| Estimates of benefit and consistency across studies | |
| What harms were identified? | |
| Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR? | |

**1a.4 OTHER SOURCE OF EVIDENCE**

*If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.*

**1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure.** A list of references without a summary is not acceptable.

Not Applicable. This measure is an outcome measure.

**1a.4.2 What process was used to identify the evidence?**

Not Applicable. This measure is an outcome measure.

**1a.4.3. Provide the citation(s) for the evidence.**

Not Applicable. This measure is an outcome measure.

## 1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

**1b.1. Briefly explain the rationale for this measure** *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

*<u>If a COMPOSITE</u> (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.*

During an Inpatient Rehabilitation Facility (IRF) stay, the goals of treatment include fostering the patient's ability to manage his or her daily activities so that the patient can complete self-care and mobility activities as independently as possible and, if feasible, return to a safe, active and productive life in a community-based setting. Given that the primary goal of rehabilitation is function improvement, IRF clinicians have traditionally assessed and documented patients' functional status at admission and discharge to calculate change in

function scores. The change in function scores represent the effectiveness of the rehabilitation care provided to patients in the rehabilitation unit or hospital.

The self-care quality measure uses standardized data elements for the collection of functional status data, which can improve communication when patients are transferred between providers. Most IRF patients receive care in an acute care hospital prior to the IRF stay, and many IRF patients receive care from another provider after the IRF stay. Use of standardized clinical data to describe a patient´s status across providers can facilitate communication across providers.

In describing the importance of functional status, the National Committee on Vital and Health Statistics Subcommittee on Health (2001) noted, "Information on functional status is becoming increasing essential for fostering healthy people and a health population. Achieving optimal health and well-being for Americans requires an understanding across the life space of the effects of people's health conditions on their ability to do basic activities and participate in life situations, in other words, their functional status."

This quality measure will inform IRF providers about opportunities to improve care in the area of function and strengthen incentives for quality improvement related to patient function.

Citation:

National Committee on Vital and Health Statistics Subcommittee on Health. Classifying and Reporting Functional Status. 2001. Retrieved from http://www.ncvhs.hhs.gov/010617rp.pdf

**1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis**. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

We provide comparisons of fiscal year 2017 and calendar year 2017 performance scores using 12 months of data, as well as scores by quarter that were conducted using the national IRF-PAI data. Performance measure scores for a more recent 12-month period (e.g., calendar year 2018) were not yet available for this analysis due to the data correction period providers have to review and correct the data. The fiscal year 2017 IRF-PAI data set includes Medicare patients discharged from IRFs between October 1, 2016 – September 30, 2017 (N=490,032) whereas the calendar year includes patients discharged between January 1, 2017 – December 31, 2017 (N=493,209) before exclusion criteria are applied.

Quality measure score distributions over two 12-month time periods:

1. Fiscal year 2017 (October 1, 2016 – September 30, 2017) (n=1,119 providers)

2. Calendar year 2017 (January 1, 2017 – December 31, 2017) (n=1,117 providers)

Quality measure score distributions by quarter between October 1, 2016 – December 31, 2017 (5 quarters):

1. Quarter 4, 2016 (n=1,103)

2. Quarter 1, 2017 (n=1,105)

3. Quarter 2, 2017 (n=1,107)

4. Quarter 3, 2017 (n=1,107)

5. Quarter 4, 2017 (n=1,096)

Quality measure score distributions over 12-months were similar between fiscal year 2017 (mean: 11.4; standard deviation: 1.7) and between calendar year 2017 (mean: 11.5; standard deviation: 1.7). Quality measure scores by decile show variations in quality measure scores across IRFs. The interquartile range for the two periods was 2.2 self-care units. Across five quarters (Q4, 2016 – Q4, 2017), mean scores increased marginally from 11.3 to 11.5 and quality measure score distributions showed variation in IRF outcomes.

12-Month Comparison

1) October 1, 2016 – September 30, 2017 (12 months)

Facilities: 1,119

Mean score: 11.4

Standard deviation: 1.7

Interquartile range: 2.2

1st decile (5.1-9.3): 8.3

2nd decile (9.4-10.0): 9.7

3rd decile (10.1-10.5): 10.3

4th decile (10.6-10.9): 10.7

5th decile (11.0-11.3): 11.2

6th decile (11.4-11.7): 11.5

7th decile (11.8-12.2): 12.0

8th decile (12.3-12.8): 12.5

9th decile (12.9-13.5): 13.1

10th decile (13.6-17.0): 14.3

Minimum: 5.1

Maximum: 17.0

2) Jan 1, 2017 – Dec 31, 2017 (12 months)

Facilities: 1,117

Mean score: 11.5

Standard deviation: 1.7

Interquartile range: 2.2

1st decile (5.4-9.4): 8.4

2nd decile (9.5-10.1): 9.9

3rd decile (10.2-10.6): 10.4

4th decile (10.7-11.0): 10.9

5th decile (11.1-11.4): 11.3

6th decile (11.5-11.8): 11.7

7th decile (11.9-12.2): 12.0

8th decile (12.3-12.8): 12.5

9th decile (12.9-13.6): 13.2

10th decile (13.7-17.5): 14.5

Minimum: 5.4

Maximum: 17.5

Quality Measure Score Distributions by Quarter

1) October 1, 2016 – December 31, 2017 (Q4, 2016)

Facilities: 1,103

Mean score: 11.3

Standard deviation: 1.9

Interquartile range: 2.5

Minimum: 2.8

Maximum: 18.9

2) January 1, 2017 – March 31, 2017 (Q1, 2017)

Facilities: 1,105

Mean score: 11.4

Standard deviation: 1.9

Interquartile range: 2.3

Minimum: 3.8

Maximum: 19.2

3) April 1, 2017 – June 30, 2017 (Q2, 2017)

Facilities: 1,107

Mean score: 11.5

Standard deviation: 1.9

Interquartile range: 2.4

Minimum: 4.4

Maximum: 17.8

4) July 1, 2017 – September 30, 2017 (Q3, 2017)

Facilities: 1,107

Mean score: 11.5

Standard deviation: 1.9

Interquartile range: 2.6

Minimum: 3.3

Maximum: 18.5

5) October 1, 2017 – December 31, 2017 (Q4, 2017)

Facilities: 1,096

Mean score: 11.5

Standard deviation: 1.9

Interquartile range: 2.4

Minimum: 1.4

Maximum: 18.3

Note: Scores are reported as units of change in self-care; Providers with < 20 stays during the 12-month testing period are excluded.

Source: RTI analysis of IRF-PAI October 2016 – December 2017 (Program reference: MV50, MV64).

**1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**

Research has shown differences in IRF patients' functional (self-care and mobility) outcomes by geographic region, facility characteristics, IRF length of stay and race/ethnicity after adjusting for key patient demographic characteristics and admission clinical status, which supports the need to monitor IRF patients' functional outcomes. We conducted a literature search to identify recent relevant studies published between 2012 and 2018 using PubMed. Among the 30 articles initially identified by the search, 15 addressed gaps in performance

for functional outcomes, and findings from these studies are summarized below. Note that the literature addresses motor functional outcomes broadly, rather than self-care or mobility specifically.

1) Variations in Functional Outcomes (Self-Care and Mobility) by Geographic Region:

We identified three studies focused on variation by geographic regions. While one study found that functional status and change in function did not vary substantially across regions (Reistetter et al., 2014), two more recent studies found significant differences in functional outcomes based on regional differences after adjusting for patient-level and facility-level characteristics (Reistetter et al., 2015; Teppala et al., 2017). Some of the variation in outcomes appear to be associated with facility-level characteristics rather than geography. Comparison of intra-class correlation coefficients from two- and three-level models showed that while the variance by facility is reduced when adjusting for random effect of hospital referral region (HRR), the reduction in the percentage of variance due to HRR is much greater when adjusting for random effect of facility. Findings suggest that there are opportunities for improvement in the area of functional status based on variations in outcomes by geographic region.

References:

Reistetter, T. A., et al. (2014). "Regional Variation in Stroke Rehabilitation Outcomes." Arch Phys Med Rehabil. 95(1), 29-38.

Reistetter T.A., et al. (2015). "Geographic and Facility Variation in Inpatient Stroke Rehabilitation: Multilevel Analysis of Functional Status." Arch Phys Med Rehabil. 96(7):1248-1254.

Srinivas Teppala, et al. (2017). "Variation in Functional Status After Hip Fracture: Facility and Regional Influence on Mobility and Self-Care." J Gerontol A Biol Sci Med Sci 72(10): 1376-1382.

2) Variations in Functional Outcomes (Self-Care and Mobility) by Facility Characteristics:

Three studies reported significant associations between facility-level characteristics and functional outcomes (Cary, et al., 2015; Graham, et al., 2013; Karmarkar, et al., 2014). Cary et al. (Cary, et al., 2015) examined variation in functional discharge scores by IRF type, ownership type, facility size as defined by number of beds, and rurality. All facility characteristics except government ownership, were associated with motor function on discharge. Using hierarchical regression modeling to estimate the association between facility characteristics and functional outcomes, the authors found that patients treated at freestanding rehabilitation hospitals, for-profit facilities, smaller facilities, and rural facilities achieved higher discharge motor scores and change in motor scores. Cary et al. noted that findings with respect to ownership type, may relate to possible selection behavior and coding practices in response to financial incentives in the Prospective Payment System.

Graham et al. (Graham, et al., 2013) examined the association between volume, as defined by average annual diagnosis facility volume for three specific diagnoses (stroke, fracture, and joint replacement) and functional outcomes. Hierarchical models showed a small, but also significant association between facility volume and functional discharge status, with the greatest effect being observed in comparing the variation between the referent and highest volume quartile.

Karmarkar et al. (2014) studied the association between IRF facility-level factors and discharge functional status of patients after stroke, accounting for patient factors. Multi-level modeling results demonstrated that although patient mix explained about 50 percent of variations in functional outcomes, facility-level factors accounted for a large part of functional outcome variations across IRFs.

Findings suggest that there are opportunities for improvement in the area of functional status based on variations in outcomes by facility characteristics.

References:

Cary, M. P., et al. (2015). "Performance-based outcomes of inpatient rehabilitation facilities treating hip fracture patients in the United States." Archives of Physical Medicine and Rehabilitation 96(5): 790-798.

Graham, J. E., et al. (2013). "Inpatient rehabilitation volume and functional outcomes in stroke, lower extremity fracture, and lower extremity joint replacement." Med Care 51(5): 404-412.

Karmarkar, A. M., et al. (2014, June). “Is Variability in Stroke Outcomes Attributable to Post-Acute Inpatient Rehabilitation Facility Factors?” AcademyHealth, San Diego, CA.

3) Variations in Functional Outcomes (Self-Care and Mobility) by IRF Length of Stay:

Several studies (O’Brien, et al., 2013; Camicia, et al., 2015; Cary, et al., 2015; Cary, et al., 2016) have shown positive associations between length of stay (LOS) and functional status at discharge, as well as functional gain. A study of IRF data spanning 2002-2007 found that since the implementation of a payment policy, LOS decreased by 1.8 days and that mean discharge FIM scores declined during the study period (O’Brien, et al., 2013).

More recent research points to more nuanced findings suggesting that the association between LOS and functional gain varies by level of impairment severity. Camicia et al.’s (Camicia et al., 2015) study of stroke patients’ functional outcomes and LOS, found longer LOS was negatively associated with functional gains of patients in the mildly impaired group, while a positive association was found among patients with moderate and severe impairments. Factors noted as possible contributors to this variation included the negative effects of hospitalization, and differences in characteristics of the various impairment groups, such as differences in age distribution, comorbidities, and functional status at admission.

References:

Camicia, M., et al. (2016). "Length of Stay at Inpatient Rehabilitation Facility and Stroke Patient Outcomes." Rehabil Nurs 41(2): 78-90.

Cary, M. P., et al. (2015). "Performance-based outcomes of inpatient rehabilitation facilities treating hip fracture patients in the United States." Archives of Physical Medicine and Rehabilitation 96(5): 790-798.

Cary, M. P., et al. (2016). "Inpatient Rehabilitation Outcomes in a National Sample of Medicare Beneficiaries With Hip Fracture." Journal of Applied Gerontology 35(1): 62-83.

O’Brien, S.R., et al. (2013). “Shorter Length of Stay is Associated with Worse Functional Outcomes for Medicare Beneficiaries With Stroke.” Phys Ther. 93(12): 1592-1602.

4) Variations in Functional Outcomes (Self-Care and Mobility) by Race and Ethnicity:

Literature focused on functional outcomes by race/ethnicity suggests lower functional outcomes for racial and ethnic minority patients, especially Black patients relative to their White counterparts (Berges, et al., 2012; Fyffe, et al., 2014; Ellis, et al., 2016; Cary, et al., 2016; Howrey, et al., 2017), though one article found no association between race and functional outcomes for patients with stroke undergoing rehabilitation (Rabadi, et al., 2012). Two studies with inconsistent findings suggest that variations in functional status or gains across race/ethnic groups may be attributable to the use of different measurement approaches (Ellis et al., 2016; Ellis et al., 2014).

References:

Berges, I-M, et al. (2012). "Recovery of Functional Status After Stroke in a Tri-Ethnic Population." PM R. 4(4): 290-295.

Cary, M. P., et al. (2016). "Inpatient Rehabilitation Outcomes in a National Sample of Medicare Beneficiaries With Hip Fracture." Journal of Applied Gerontology 35(1): 62-83.

Ellis, C., et al. (2014). "Racial/Ethnic Differences in Poststroke Rehabilitation Outcomes." Stroke Research and Treatment.

Ellis, C., et al. (2016). "Racial Differences in Poststroke Rehabilitation Utilization and Functional Outcomes." Arch Phys Med Rehabil. 96: 84-90.

Fyffe, D.C., et al. (2014). "Racial and Ethnic Disparities in Functioning at Discharge and Follow-Up Among Patients With Motor Complete Spinal Cord Injury." Am J Phys Med Rehabil 95: 2140-51.

Howrey, B.T., et al. (2017). "Trajectories of Functional Change after Inpatient Rehabilitation for Traumatic Brain Injury." Arch Phys Med Rehabil 98(8): 1606-1613.

Rabadi, M. H., et al. (2012). "Does race influence functional outcomes in patients with acute stroke undergoing inpatient rehabilitation?" Am J Phys Med Rehabil 91(5): 375-382; quiz 383-376.

**1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

We used the 2017 national IRF-PAI data set, which includes all Medicare patients discharged from IRFs in calendar year 2017, to examine whether there may be disparities in care for population groups related to this measure. Disparities for certain population groups would indicate gaps in care and opportunities for improvement. The 2017 national IRF-PAI data set included 1,129 IRFs who discharged 493,209 patients in 2017.

We address the issue of disparities for this measure by examining whether there are differences in functional outcomes for population groups that may reflect experience disparities in care, such as for population groups with social risk factors.

We examined whether 5 social risk factors were associated with change in self-care scores, after risk adjustment: 1) dual eligibility (patient-level variable); 2) race/ethnicity (patient-level variable); 3) living alone (patient-level variable); 4) urbanicity based on the patient's residence (community-level variable); and 5) socioeconomic status (SES) (community-level variable). Details about how we obtained and calculated this disparities data is available in Sections 1.2 and 1.8 of the Testing form.

We conducted the following analyses to examine the effect of the 5 social risk factors:

1) We calculated the percentage of stays for each social risk factor population group;
2) We calculated the observed change in self-care score for each social risk factor population group;
3) We added indicators for each social risk factor group to our risk adjustment model and estimated the coefficients for each group (relative to the reference group) in the model;
4) We examined the indicators for each social risk factor over time by quarter in our risk adjustment model to examine whether there may be trends for population groups.

Below is a summary of these analyses and results. For more information on disparities in change in self-care related to dual eligibility, race/ethnicity, living alone, urbanicity and SES, please refer to the risk adjustment analyses in the Testing form. Tables and graphics are able to be inserted into the NQF Testing form, unlike this Measure Information form, so we direct readers to Section 2b3.4b of the Testing form for the results presented below in a more readable format (Tables 13, 14, and 15 specifically).

1) The Distribution of Social Risk Factor Patient Population Groups:

We found that 12.2% of patients were dually-eligible with full Medicaid benefits, 79.4% of patients were white, and 29.7% were living alone. We also found that 83.8% of IRF patients lived in urban areas. The lowest quartile of AHRQ SES index ranged from 27.9 - 49.5; the highest quartile ranged from 55.3 – 75.7.

2) Observed Change in Self-Care Score by Social Risk Factor:

The mean unadjusted (observed) change in self-care score varied slightly by dual eligibility status, race, and living alone status. Dual eligible patients with full Medicaid benefits had on average 11.0 units of change in self-care while patients who were dual eligible without full Medicaid benefits or who were non-dual eligible had more change in self-care (12.0 and 11.6 units, respectively). For race, the highest mean change in self-care was found among patients who were white (11.6 units of change), multiracial (11.5 units of change), or Native American or Alaskan Native (11.4 units of change) whereas the lowest was among patients who were Asian (10.4 units of change). Patients who were living alone prior to their hospitalization had on average 12.0 units of

change in self-care whereas those not living alone had 11.3 units of change in self-care. The mean unadjusted (observed) change in self-care scores were similar across Hispanic ethnicity, urbanicity, and SES.

3) Estimated Effect (Coefficient Values) for Each Social Risk Factor (Full Year)

Each social risk factor was then added to our Generalized Linear regression model to get estimated regression coefficients which represent the effect of each individual factor on change in self-care relative to the refence group. The dependent variable was the change in self-care score for each patient, calculated as the difference between the discharge self-care score and admission self-care score. For example, a coefficient value of -0.5 for Black patients would be interpreted to mean that, on average, these patients had a change in self-care score that was 0.5 self-care units less than White patients (the reference group).

Lower self-care change scores were observed and significant for dual eligibility patients with full Medicaid benefits compared to non-duals. Black patients, Asian patients, and patients of Native Hawaiian or Pacific Islander descent also had lower self-care changes scores compared to White patients. Hispanic patients, on the other hand, had higher self-care change scores than non-Hispanic patients. Other population groups with higher self-care changes scores included patients who lived alone compared to patients who did not prior to their hospitalization, and patients residing in AHRQ SES Index quartiles 1-3 (i.e., lower SES areas) than patients residing in AHRQ SES Index quartile 4 (i.e., the highest SES areas).

4) Estimated Coefficient Values for Each Social Risk Factor (by Quarter)

The 2017 analysis described above examining each social risk factor's effect on change in self-care was then performed by quarter to examine possible trends over time (Q1, 2017 – Q4, 2017). The patients included in each quarter and detailed results are provided below.

The differences observed with the full calendar year 2017 data were generally found to be consistent by quarter. The population groups with slightly lower self-care changes scores or higher self-care change scores continued to show these differences. Specifically, the coefficient value for dual eligibility patients with full Medicaid benefits ranged from -0.3068 to -0.4479 depending on the quarter compared to the self-care change scores for non-dual eligible patients. On average, Black patients (coeff. range = -0.5013 to -0.6370), and patients of Native Hawaiian or Pacific Islander descent (coeff. range = -0.1762 to -0.6648) had slightly significantly lower self-care change scores than White patients. For Asian patients, a trend was observed of less improvement compared to White patients across the 4 quarters (coeff. range -0.0760 to -0.7107).

For the population groups with higher self-care changes scores, quarterly results indicate the trend remained for patients who lived alone compared to patients who did not prior to their hospitalization (coeff. range = 0.4267 to 0.5248). For patients residing in AHRQ SES Index quartiles 1-3 (i.e., lower SES areas) we observe higher self-care change scores in all quarters compared to the AHRQ SES Index quartile 4 (i.e., the highest SES areas). Specifically, the lowest SES group quartile 1 had the highest coefficient estimates (coeff. range = 0.5011 to 0.6138) compared to the highest SES group. As socioeconomic status increases for those living in SES quartiles 2 and 3, the coefficient estimates are smaller (less effect). The coefficients ranged from 0.3810 to 0.4103 for SES quartile 2 and 0.2483 to 0.3345 for SES quartile 3.

Our testing of social risk factors and their relationships to patients' change in self-care scores indicate that some factors (full dual eligibility, Black, Asian or Native Hawaiian race) were tied to slightly lower self-care change scores while others (lower SES, living alone, Hispanic ethnicity) were tied to higher self-care change scores. Though the effects on lower changes in self-care scores were small, we believe that continued monitoring of potential disparities in functional outcomes is critical.

Breakdown of patients discharged within each quarter:

Jan 1 – Mar 31, 2017 (Q1 2017) = 107,464

Apr 1 – Jun 30, 2017 (Q2 2017) = 107,611

Jul 1 – Sept 30, 2017 (Q3 2017) = 104,831

Oct 1 – Dec 31, 2017 (Q4 2017) = 108,286

Dual Eligibility (reference = Non-dual)

Dual with full Medicaid

- Q1 2017: estimate = -0.3876; SE = 0.06; p-value <.0001
- Q2 2017: estimate = -0.3589; SE = 0.06; p-value <.0001
- Q3 2017: estimate = -0.4479; SE = 0.06; p-value <.0001
- Q4 2017: estimate = -0.3068; SE = 0.06; p-value <.0001

Dual without full Medicaid

- Q1 2017: estimate = 0.3601; SE = 0.08; p-value <.0001
- Q2 2017: estimate = 0.0924; SE = 0.08; p-value = 0.2224
- Q3 2017: estimate = 0.2826; SE = 0.08; p-value = 0.0002
- Q4 2017: estimate = 0.2444; SE = 0.08; p-value = 0.0016

Race/Ethnicity (reference = White)

Black

- Q1 2017: estimate = -0.6370; SE = 0.06; p-value <.0001
- Q2 2017: estimate = -0.5364; SE = 0.06; p-value <.0001
- Q3 2017: estimate = -0.5423; SE = 0.06; p-value <.0001
- Q4 2017: estimate = -0.5013; SE = 0.06; p-value <.0001

Asian

- Q1 2017: estimate = -0.0760; SE = 0.14; p-value = 0.5909
- Q2 2017: estimate = -0.4846; SE = 0.14; p-value = 0.0006
- Q3 2017: estimate = -0.6475; SE = 0.14; p-value <.0001
- Q4 2017: estimate = -0.7107; SE = 0.14; p-value <.0001

American Indian or Alaska Native

- Q1 2017: estimate = -0.3632; SE = 0.31; p-value = 0.2482
- Q2 2017: estimate = -0.2837; SE = 0.31; p-value = 0.3528
- Q3 2017: estimate = 0.1890; SE = 0.31; p-value = 0.5361
- Q4 2017: estimate = 0.0457; SE = 0.32; p-value = 0.8859

Native Hawaiian or Pacific Islander

- Q1 2017: estimate = -0.1762; SE = 0.29; p-value = 0.5456
- Q2 2017: estimate = -0.6648; SE = 0.28; p-value = 0.0191
- Q3 2017: estimate = -0.6505; SE = 0.29; p-value = 0.0250
- Q4 2017: estimate = -0.1952; SE = 0.29; p-value = 0.5026

Multiracial

- Q1 2017: estimate = 0.8234; SE = 0.69; p-value = 0.2315
- Q2 2017: estimate = -0.8240; SE = 0.79; p-value = 0.2966
- Q3 2017: estimate = 0.2526; SE = 0.73; p-value = 0.7310
- Q4 2017: estimate = -0.0440; SE = 0.69; p-value = 0.9495

Hispanic Ethnicity

- Q1 2017: estimate = -0.1185; SE = 0.13; p-value = 0.3445

- Q2 2017: estimate = 0.4234; SE = 0.12; p-value = 0.0005
- Q3 2017: estimate = 0.2362; SE = 0.12; p-value = 0.0528
- Q4 2017: estimate = 0.1272; SE = 0.12; p-value = 0.2976

Living Alone

- Q1 2017: estimate = 0.4267; SE = 0.04; p-value <.0001
- Q2 2017: estimate = 0.4347; SE = 0.04; p-value <.0001
- Q3 2017: estimate = 0.5034; SE = 0.04; p-value <.0001
- Q4 2017: estimate = 0.5248; SE = 0.04; p-value <.0001

Urbanicity (reference = Urban)

Rural

- Q1 2017: estimate = 0.0575; SE = 0.09; p-value = 0.5194
- Q2 2017: estimate = -0.1315; SE = 0.09; p-value = 0.1333
- Q3 2017: estimate = -0.1101; SE = 0.09; p-value = 0.2184
- Q4 2017: estimate = -0.0025; SE = 0.09; p-value = 0.9770

Suburban

- Q1 2017: estimate = 0.0584; SE = 0.06; p-value = 0.3028
- Q2 2017: estimate = -0.0001; SE = 0.06; p-value = 0.9991
- Q3 2017: estimate = -0.0347; SE = 0.06; p-value = 0.5489
- Q4 2017: estimate = 0.0905; SE = 0.06; p-value = 0.1122

AHRQ SES Index* (reference = Quartile 4)

Quartile 1

- Q1 2017: estimate = 0.5714; SE = 0.05; p-value <.0001
- Q2 2017: estimate = 0.6138; SE = 0.05; p-value <.0001
- Q3 2017: estimate = 0.5011; SE = 0.05; p-value <.0001
- Q4 2017: estimate = 0.5264; SE = 0.05; p-value <.0001

Quartile 2

- Q1 2017: estimate = 0.4074; SE = 0.05; p-value <.0001
- Q2 2017: estimate = 0.3928; SE = 0.05; p-value <.0001
- Q3 2017: estimate = 0.4103; SE = 0.05; p-value <.0001
- Q4 2017: estimate = 0.3810; SE = 0.05; p-value <.0001

Quartile 3

- Q1 2017: estimate = 0.3345; SE = 0.05; p-value <.0001
- Q2 2017: estimate = 0.2483; SE = 0.05; p-value <.0001
- Q3 2017: estimate = 0.3146; SE = 0.05; p-value <.0001
- Q4 2017: estimate = 0.2757; SE = 0.05; p-value <.0001

* based on patient residence. AHRQ = Agency for Healthcare Research.

Note: SE=Standard error; Patient-level exclusion criteria applied. Data missing for Race, Urbanicity, and AHRQ SES Index not displayed.

Source: RTI analysis of IRF-PAI, January – December 2017. (Program reference: LP65)

**1b.5. If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4**

We conducted a literature search to identify recent relevant manuscripts published between 2012 and 2018 using PubMed that examined disparities in functional outcomes among IRF patients. We identified 7 studies that focused on differences in outcomes by race/ethnicity group. Findings from these studies are summarized below. Note that the literature addresses motor functional outcomes broadly, rather than self-care or mobility specifically.

Literature focused on functional outcomes by race/ethnicity suggests lower functional outcomes for racial and ethnic minority patients, especially Black patients relative to their White counterparts (Berges, et al., 2012; Fyffe, et al., 2014; Ellis, et al., 2016; Cary, et al., 2016; Howrey, et al., 2017), though one article found no association between race and functional outcomes for patients with stroke undergoing rehabilitation (Rabadi, et al., 2012). Two studies with inconsistent findings suggest that variations in functional status or gains across race/ethnic groups may be attributable to the use of different measurement approaches (Ellis et al., 2016; Ellis et al., 2014).

References:

Berges, I-M, et al. (2012). "Recovery of Functional Status After Stroke in a Tri-Ethnic Population." PM R. 4(4): 290-295.

Cary, M. P., et al. (2016). "Inpatient Rehabilitation Outcomes in a National Sample of Medicare Beneficiaries With Hip Fracture." Journal of Applied Gerontology 35(1): 62-83.

Ellis, C., et al. (2014). "Racial/Ethnic Differences in Poststroke Rehabilitation Outcomes." Stroke Research and Treatment.

Ellis, C., et al. (2016). "Racial Differences in Poststroke Rehabilitation Utilization and Functional Outcomes." Arch Phys Med Rehabil. 96: 84-90.

Fyffe, D.C., et al. (2014). "Racial and Ethnic Disparities in Functioning at Discharge and Follow-Up Among Patients With Motor Complete Spinal Cord Injury." Am J Phys Med Rehabil 95: 2140-51.

Howrey, B.T., et al. (2017). "Trajectories of Functional Change after Inpatient Rehabilitation for Traumatic Brain Injury." Arch Phys Med Rehabil 98(8): 1606-1613.

Rabadi, M. H., et al. (2012). "Does race influence functional outcomes in patients with acute stroke undergoing inpatient rehabilitation?" Am J Phys Med Rehabil 91(5): 375-382; quiz 383-376.

Summary of each study:

Berges, I-M., et al. (2012). "Recovery of Functional Status After Stroke in a Tri-Ethnic Population." PM R. 4(4): 290-295.

- Examined differences in functional status for White, Black and Hispanic stroke patients from time of admission to an IRF up to 12 months after discharge.
- Study design: longitudinal study of stroke patient data (n = 990) from the Stroke Recovery in Underserved Populations database (2005-2006). Patients were age 55 or older and were interviewed at 4 points: admission to IRF, discharge, 3 months after discharge, 12 months after discharge.
- Race and ethnicity were amongst the significant predictors of total FIM scores.
- Differences between the groups differed across the various time periods: during rehabilitation, both Black and Hispanic function admission scores were slightly higher than those of their White counterparts and functional gains were similar; however, at the 3-month follow-up, scores for Black and Hispanic patients were lower than those of White patients, and at the 12-month follow-up, only Hispanic patients continued to have significantly lower scores than White patients.

- Study findings suggest that variations in recovery across race/ethnic groups may have more to do with post-rehabilitation factors.

Cary, M. P., et al. (2016). "Inpatient Rehabilitation Outcomes in a National Sample of Medicare Beneficiaries With Hip Fracture." Journal of Applied Gerontology 35(1): 62-83.

- Black, Hispanic, and Other racial/ethnic patients had lower FIM scores at discharge compared to White patients; FIM discharge scores of Asian patients were similar to those of White patients.
- It is important to note that the regression model that included only "predisposing variables" (age, sex, and race) explained only 9% of the variance.

Ellis, C., et al. (2016). "Racial Differences in Poststroke Rehabilitation Utilization and Functional Outcomes." Arch Phys Med Rehabil. 96: 84-90.

- Examined racial differences in post-stroke rehabilitation utilization and functional outcomes.
- Study design: A follow-up study of stroke survivors 45 years or older seen for stroke care from October 1, 2008, to September 30, 2009 at a stroke center in South Carolina.
- Black patients had lower levels of overall functional independence than did White patients (8.0 vs 10.5; P<.05).
- "Three key findings emerged from the study: (1) blacks experienced higher levels of impairment at stroke onset than did whites, (2) blacks reported lower levels of functional independence at 1 year poststroke onset, and (3) blacks reported lower levels of functional independence and driving independence despite a lack of racial differences in rehabilitation utilization."
- Note that part of inconsistency in findings regarding racial disparities in functional outcomes can be attributed to use of different measurement approaches and variation of settings.

Ellis, C., et al. (2014). "Racial/Ethnic Differences in Poststroke Rehabilitation Outcomes." Stroke Research and Treatment.

- Examined racial and ethnic differences in poststroke rehabilitation outcomes.
- Study design: Literature review of articles on stroke, rehabilitation, and racial-ethnic patterns of disease over a 10-year period (2003–2012) and focused on rehabilitation outcomes and the race or ethnicity of at least two groups.
- Majority of the studies found that racial/ethnic minorities were less likely to achieve equivalent functional improvement following rehabilitation. Blacks were more likely to experience lower FIM gain or change scores (range: 1–60%) and more likely to have lower efficiency scores (range: 5–16%) than Whites.
- Here to, note of variability of study approaches and resulting difficulty of drawing conclusions from the findings.

Fyffe, D.C., et al. (2014). "Racial and Ethnic Disparities in Functioning at Discharge and Follow-Up Among Patients With Motor Complete Spinal Cord Injury." Am J Phys Med Rehabil 95: 2140-51.

- Examined racial and ethnic differences in self-care and mobility outcomes for persons with a motor complete, traumatic spinal cord injury (SCI) at discharge and 1-year follow-up.
- Study design: retrospective cohort study using patient data from the Spinal Cord Injury Model Systems (SCIMS) database for patients enrolled in the SCIMS between 2000-2011 (n=1766).
- At discharge, non-Hispanic black participants with tetraplegia and paraplegia had significantly poorer gains in FIM self-care and mobility scores relative to non-Hispanic white and Hispanic participants. [Discussion notes that the difference is small.]
- At 1-year follow-up, similar FIM self-care and mobility change scores were found across racial and ethnic groups within each neurologic category.

Howrey, B.T., et al. (2017). "Trajectories of Functional Change after Inpatient Rehabilitation for Traumatic Brain Injury." Arch Phys Med Rehabil 98(8): 1606-1613.

- Examined trajectories of functional recovery after rehabilitation for TBI.
- Study design: prospective study of IRF TBI patients from 2002 to 2010 who also had post-discharge measurements of functional independence (n = 16,583) using UDS data.
- Being of a racial/ethnic minority was associated with membership in the low motor trajectory

Rabadi, M. H., et al. (2012). "Does race influence functional outcomes in patients with acute stroke undergoing inpatient rehabilitation?" Am J Phys Med Rehabil 91(5): 375-382; quiz 383-376.

- Examined relationship between race and functional outcomes on stroke patients receiving facility-based rehabilitation.
- Study design: 2-year prospective study of patients admitted to an acute stroke rehabilitation unit within 30 days after an acute stroke (n=670).
- The primary and secondary functional rehabilitation outcomes were similar for all four groups after similar intensity of therapy (3.5 hours/day).
- Found no significant association between race and functional outcomes.

## 2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

**2a.1. Specifications** The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

**De.5. Subject/Topic Area** *(check all the areas that apply):*

Musculoskeletal, Musculoskeletal : Falls and Traumatic Injury, Neurology : Brain Injury, Neurology : Stroke/Transient Ischemic Attack (TIA)

**De.6. Non-Condition Specific***(check all the areas that apply):*

Health and Functional Status : Change

**De.7. Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*

Populations at Risk : Individuals with multiple chronic conditions

**S.1. Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Functional-Measures-.html

**S.2a. If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)

This is not an eMeasure **Attachment:**

**S.2b. Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*

Attachment **Attachment:** Change_in_Self-Care_NQF_2633_Risk_Adj_Model_01-07-2019.xlsx

**S.2c.** Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Attachment **Attachment:** Final_IRF-PAI_Version_3.0_-_Effective_October_1_2019_-FY2020-.pdf

**S.2d.** Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Clinician

**S.3.1. <u>For maintenance of endorsement:</u>** Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

Yes

**S.3.2. <u>For maintenance of endorsement,</u>** please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

We have made several changes to the specifications, including updates to the exclusion criteria, risk adjustors, and measure calculation algorithm since the most recent annual update:

(1) Exclusion criteria: We are removing "discharged to another IRF" as an exclusion criterion from the incomplete stay definition. Rationale: The removal of this criterion means that the definition of an "incomplete stay" for this measure is aligned with other post-acute care function quality measures. When a patient is discharged to another IRF, the discharge would not typically be urgent, so gathering discharge functional assessment data for these patients is feasible.

(2) Risk-Adjustors: We have updated the covariates included in the risk adjustment model by removing several comorbidities and adding low body mass index (BMI) and several comorbidities. Rationale: When examining the risk adjustment model using the 12-month national IRF-PAI data, we found that some comorbidities were no longer significant predictors of change in self-care or the association between the comorbidity and functional outcomes was no longer consistent with the literature or clinical expectations. Following a literature review, we tested additional candidate risk adjusters. We added low BMI and several comorbidities (hierarchical condition category groups) to the regression model based on the magnitude of the coefficient that suggested the comorbidity was an important factor associated with functional outcomes among IRF patients. Adding these risk adjustors to the model will not add provider burden, because the data are already collected via the IRF-PAI.

(3) Measure Calculation: The risk-adjustment procedure for this measure involves comparing patients' observed change in self-care scores with their expected change in self-care scores. We are revising this part of the measure calculation. The prior approach used the ratio of the observed to expected values and the ratio was multiplied by the national mean. The new approach uses the difference between the observed and expected values, and the difference value is added to the national mean. Rationale: We have developed an application of this measure for skilled nursing facilities (SNFs) and use the difference approach for the SNF measure given the potential for more variation in the observed and expected values due to a more heterogeneous SNF population. We are now updating this IRF functional outcome measure to use the difference approach so the IRF and SNF measure calculations are aligned. Our testing of the two approaches (ratio and difference approaches) with national IRF data showed no meaningful difference in the facility mean and median quality measure scores.

**S.4. Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*

*<u>IF an OUTCOME MEASURE</u>, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).*

The measure does not have a simple form for the numerator and denominator. This measure estimates the risk-adjusted change in self-care score between admission and discharge among Inpatient Rehabilitation Facility (IRF) Medicare Part A and Medicare Advantage patients age 21 or older. The change in self-care score is calculated as the difference between the discharge self-care score and the admission self-care score.

**S.5. Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

*IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).*

Seven self-care activities are each scored based on a patient´s ability to complete the activity. The scores for the seven activities are summed to obtain a self-care score at the time of admission and at the time of discharge. The change in self-care is the difference between the discharge self-care score and the admission self-care score.

The 7 self-care items are:

GG0130A. Eating

GG0130B. Oral hygiene

GG0130C. Toileting hygiene

GG0130E. Shower/bathe self

GG0130F. Upper body dressing

GG0130G. Lower body dressing

GG0130H. Putting on/taking off footwear

Each patient´s ability to complete each self-care activity (item) is rated by a clinician using the following 6-level rating scale:

level 06 - Independent

level 05 - Setup or clean up assistance

level 04 - Supervision or touching assistance

level 03 - Partial/moderate assistance

level 02 - Substantial/maximal assistance

level 01 - Dependent

If the patient did not attempt the activity, the reason that the activity did not occur is reported as:

07 = Patient refused

09 = Not applicable

10 = Not attempted due to environmental limitations

88 = Not attempted due to medical condition or safety concerns.

The performance period is 12 months for reporting on CMS's IRF Compare website.

**S.6. Denominator Statement** *(Brief, narrative description of the target population being measured)*

The denominator is the number of Inpatient Rehabilitation Facility Medicare patient stays, except those that meet the exclusion criteria.

**S.7. Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets –*

*Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

*IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).*

The denominator is the number of Inpatient Rehabilitation Facility Medicare Part A and Medicare Advantage patient stays, except those that meet the exclusion criteria.

**S.8. Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*

This quality measure has six patient-level exclusion criteria:

1) Patients with incomplete stays.

Rationale: When a patient has an incomplete stay, for example, the patients leave urgently due to a medical emergency, it can be challenging to gather accurate discharge functional status data. Patients with incomplete stays include patients who are unexpectedly discharged to an acute care setting (Short-stay Acute Hospital, Critical Access Hospital, Inpatient Psychiatric Facility, or Long-term Care Hospital); patients who die or leave an Inpatient Rehabilitation Facility (IRF) against medical advice; and patients with a length of stay less than 3 days.

2) Patients who are independent with all self-care activities at the time of admission.

Rationale: Patients who are independent with all the self-care items at the time of admission are assigned the highest score on all the self-care items, and thus, would not be able to show functional improvement on this same set of items at discharge.

3) Patients with the following medical conditions on admission: coma; persistent vegetative state; complete quadriplegia; locked-in syndrome; or severe anoxic brain damage, cerebral edema or compression of the brain.

Rationale: These patients are excluded because they may have limited or less predictable self-care improvement with the selected self-care items.

4) Patients younger than age 21.

Rationale: There is only limited evidence published about functional outcomes for individuals with Medicare who are younger than 21.

5) Patients discharged to Hospice.

Rationale: Patient goals may change during the IRF stay, and functional improvement may no longer be a goal for a patient discharged to hospice.

6) Patients who are not Medicare Part A and Medicare Advantage beneficiaries.

Rationale: IRF-PAI data for patients not covered by the Medicare program are not submitted to the Centers for Medicare and Medicaid Services.

Facility-level quality measure exclusion: For IRFs with fewer than 20 patient stays, data for this quality measure are not publicly reported.

**S.9. Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

The following items are used to identify which patients are excluded from the quality measure calculations.

These data elements are included on the current version of the Inpatient Rehabilitation Facility Patient Assessment Instrument (IRF-PAI), which is available at: http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/InpatientRehabFacPPS/IRFPAI.html

It can be challenging to gather accurate discharge functional status data for patients who experience incomplete stays. Patients with incomplete stays include patients who are unexpectedly discharged to an

acute care setting (Short-stay Acute Hospital, Critical Access Hospital, Inpatient Psychiatric Facility, or Long-term Care Hospital); patients discharged to a hospice; patients who die or leave an Inpatient Rehabilitation Facility (IRF) against medical advice; and patients with a length of stay less than 3 days.

Items used to identify these patient records:

1) Patients with incomplete stays.

Patients with a length of stay less than 3 days: We calculate length of stay using the following items on the IRF-PAI.

Item 12. Admission Date.

Item 40. Discharge Date.

Length of stay is calculated as the Discharge Date minus the Admission Date (Discharge Date - Admission Date). Patient records with a length of stay of less than 3 days are excluded.

Item 41. Patient discharged against medical advice. This item is used to identify patients discharged against medical advice.

Patient records with a response of “Yes = 1" are excluded.

Item 44C. Was the patient discharged alive? This item is used to identify patients who died during the IRF stay.

Patient records with a response of “No=0" are excluded.

44D. Patient’s discharge destination/living setting. This item is used to identify patients with an incomplete stay.

Short-term General Hospital = 02

Long-Term Care Hospital = 63

Inpatient Psychiatric Facility = 65

Critical Access Hospital = 66.

2) Patients who are independent with all self-care activities at the time of admission: Patients who are independent with all the self-care items at the time of admission are assigned the highest score on all the self-care items, and thus, would not be able to show functional improvement (i.e., a higher score)on this same set of items at discharge.

Self-care items

GG0130A. Eating = 06, and

GG0130B. Oral hygiene = 06, and

GG0130C. Toileting hygiene = 06, and

GG0130E. Shower/bathe self = 06, and

GG0130F. Upper body dressing = 06, and

GG0130G. Lower body dressing = 06, and

GG0130H. Putting on/taking off footwear = 06.

3) Patients with the following medical conditions on admission: coma; persistent vegetative state; complete quadriplegia; and locked-in syndrome; and severe anoxic brain damage, cerebral edema or compression of the brain.

The following items will be used to identify patients with these conditions:

21A. Impairment Group.

0004.1221 - Spinal Cord Dysfunction, Non-Traumatic: Quadriplegia Complete, C1-C4

0004.1222 - Spinal Cord Dysfunction, Non-Traumatic: Quadriplegia Complete, C5-C8

0004.2221 - Spinal Cord Dysfunction, Traumatic: Quadriplegia Complete, C1-C4

0004.2222 - Spinal Cord Dysfunction, Traumatic: Quadriplegia Complete, C5-C8

22. Etiologic Diagnosis.

This item is used to determine a patient´s etiologic problem that led to the condition for which the patient is receiving rehabilitation. The following Hierarchical Condition Categories (HCCs) and International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) codes will be used to identify and exclude the records of patients with these conditions:

HCC 80. Coma, Brain Compression/Anoxic Damage

ICD-10-CM. G82.51 Quadriplegia, C1-C4 complete

ICD-10-CM. G82.53 Quadriplegia, C5-C7 complete

ICD-10-CM. S14.11xx Quadriplegia, Complete lesion at Cx(1-8) level of cervical spinal cord, initial encounter or subsequent encounter, or sequela

ICD-10-CM. G83.5. Locked-in state

24. Comorbid Conditions.

This item is used to exclude selected comorbidities. The following Hierarchical Condition Categories (HCCs) and International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) codes will be used to identify and exclude the records of patients with these conditions:

HCC 80. Coma, Brain Compression/Anoxic Damage

ICD-10-CM. G82.51 Quadriplegia, C1-C4 complete

ICD-10-CM. G82.53 Quadriplegia, C5-C7 complete

ICD-10-CM. S14.11xx Quadriplegia, Complete lesion at Cx(1-8) level of cervical spinal cord, initial encounter or subsequent encounter, or sequela

ICD-10-CM. G83.5. Locked-in state

4) Patients younger than age 21.

These items are used to calculate age, and patients who are younger than 21 years of age at the time of admission are excluded.

6. Birth Date

12. Admission Date

Age is calculated as the Admission Date minus the Birth Date (Admission Date - Birth Date). Patients younger than 21 are excluded.

5) Patients discharged to hospice

44D. Patient’s discharge destination/living setting.

This item is used to identify patients discharged to hospice. The following responses are used:

Hospice (home) = 50

Hospice (institutional facility) = 51

6) Patients who are not Medicare Part A and Medicare Advantage beneficiaries.

The following items are used to identify and exclude the records of patients who are not Medicare Part A and Medicare Advantage beneficiaries:

20A. Primary Source = 99 - Not Listed AND

20B. Secondary Source = 99 - Not Listed

**S.10. Stratification Information** *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*

Not applicable. This measure does not use stratification for risk-adjustment.

**S.11. Risk Adjustment Type** (Select type. Provide specifications for risk stratification in measure testing attachment)

Statistical risk model

If other:

**S.12. Type of score:**

Continuous variable, e.g. average

If other:

**S.13. Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*

Better quality = Higher score

**S.14. Calculation Algorithm/Measure Logic** (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*

We provide the detailed calculation algorithm in an attachment entitled "IRF Detailed Function QM Specifications 2633 01-07-2019" included in the Appendix.

The detailed calculation algorithm is provided to the public in the document entitled IRF Measure Calculations and Reporting User's Manual. The current version of this document is available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/IRF-Quality-Reporting/IRF-Quality-Reporting-Program-Measures-Information-.html

The following are the key steps used to calculate the measure:

1) Sum the scores of the admission self-care items to create an admission self-care score for each patient, after 'activity not attempted' codes (07. Patient refused, 09. Not applicable, 10. Not attempted due to environmental limitations, and 88. Not attempted due to medical condition or safety concerns), skip codes ('^') and missing data ('-') are recoded. (range: 7 to 42).
2) Sum the scores of the discharge self-care items to create a discharge self-care score for each patient, after 'activity not attempted' codes (07. Patient refused, 09. Not applicable, 10. Not attempted due to environmental limitations, and 88. Not attempted due to medical condition or safety concerns), skip codes ('^') and missing data ('-') are recoded. (range: 7 to 42).
3) Identify the records of patients who meet the exclusion criteria and exclude them from analyses.
4) Calculate the difference between the admission self-care score (from step 1) and the discharge self-care score (from step 2) for each patient to create a change in self-care score for each patient.
5) Calculate an expected change in self-care score for each patient using regression coefficients from national data and each patient's admission characteristics (risk adjustors).
6) Calculate an average change in self-care score for each IRF. This is the facility-level observed change in self-care score.
7) Calculate an average expected change in self-care score for each IRF. This is the facility-level expected change in self-care score.

8) Subtract the facility-level expected change score from the facility -level observed change score to determine the difference in scores (difference value). A difference value that is 0 indicates the observed and expected scores are equal. An observed minus expected difference value that is higher than 0 (positive value) indicates that the observed change score is greater (better) than the expected change score. An observed minus expected difference value that is less than 0 (negative value) indicates that the observed change score is lower (worse) than the expected change score.

9) Add the national average change in self-care score to each IRF's difference value (from step 8). This is the risk-adjusted mean change in self-care score.

Each patient´s ability to complete each self-care activity (item) is rated by a clinician using the following 6-level rating scale:

level 06 - Independent

level 05 - Setup or clean up assistance

level 04 - Supervision or touching assistance

level 03 - Partial/moderate assistance

level 02 - Substantial/maximal assistance

level 01 - Dependent

The 7 self-care items are:

GG0130A. Eating

GG0130B. Oral hygiene

GG0130C. Toileting hygiene

GG0130E. Shower/bathe self

GG0130F. Upper body dressing

GG0130G. Lower body dressing

GG0130H. Putting on/taking off footwear

**S.15. Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.

Not applicable. This measure uses IRF-PAI data for all Medicare patients treated by IRFs for the performance period. There is no sampling. This is an instrument-based measure that relies on clinician-reported data, therefore proxy responses are not relevant.

**S.16. Survey/Patient-reported data** *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*

Specify calculation of response rates to be reported with performance measure results.

Not applicable. This measure uses clinician-reported data.

**S.17. Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

*If other, please describe in S.18.*

Instrument-Based Data

**S.18. Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*

IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.

Inpatient Rehabilitation Facility Patient Assessment Instrument (IRF-PAI)

**S.19. Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

Available at measure-specific web page URL identified in S.1

**S.20. Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*

Facility

**S.21. Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Post-Acute Care

If other:

**S.22. <u>COMPOSITE Performance Measure</u>** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

Not applicable. This is not a composite measure.

**2. Validity – See attached Measure Testing Submission Form**

NQF_IRF_Self-Care_Change_Testing_Final-636794380523984218.docx,2633_nqf_testing_4-22-2019.docx

**2.1 <u>For maintenance of endorsement</u>**

*Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.*

Yes

**2.2 <u>For maintenance of endorsement</u>**

*Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.*

Yes

**2.3 <u>For maintenance of endorsement</u>**

*Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.*

Yes - Updated information is included

## Measure Testing (subcriteria 2a2, 2b1-2b6)

**Measure Number** (*if previously endorsed*)**:** 2633
**Measure Title**: Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Change in Self-Care Score for Medical Rehabilitation Patients
**Date of Submission**: 1/7/2019

**<u>Type of Measure:</u>**

| | |
|---|---|
| ☒ Outcome (*including PRO-PM*) | ☐ Composite – ***STOP – use composite testing form*** |
| ☐ Intermediate Clinical Outcome | ☐ Cost/resource |
| ☐ Process *(including Appropriate Use)* | ☐ Efficiency |
| ☐ Structure | |

**1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE**

*Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.*

**1.1. What type of data was used for testing**? (*Check all the sources of data identified in the measure specifications and data used for testing the measure*. *Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

| **Measure Specified to Use Data From: (*must be consistent with data sources entered in S.17*)** | **Measure Tested with Data From:** |
|---|---|
| ☐ abstracted from paper record | ☐ abstracted from paper record |
| ☐ claims | ☐ claims |
| ☐ registry | ☐ registry |
| ☐ abstracted from electronic health record | ☐ abstracted from electronic health record |
| ☐ eMeasure (HQMF) implemented in EHRs | ☐ eMeasure (HQMF) implemented in EHRs |
| ☒ other: Inpatient Rehabilitation Facility Patient Assessment Instrument (IRF-PAI) | ☒ other: Inpatient Rehabilitation Facility Patient Assessment Instrument (IRF-PAI) |

**1.2. If an existing dataset was used, identify the specific dataset** (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home . health OASIS, clinical registry*).

The primary dataset used for calculating this performance measure was the National IRF-PAI data. A copy of the IRF-PAI can be found on the following website: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/IRF-Quality-Reporting/IRF-PAI-and-IRF-QRP-Manual.html

We used two additional data sources for measure testing only to provide facility and patient-level characteristics not available in the IRF-PAI. These sources are not used for quality measure calculation:

For analyses that involved facility characteristics, we used the Provider of Service file:

- **Provider of Services Current Files (POS File):** We used the POS file to describe the characteristics of IRFs, such as census region, ownership type, and rurality, reported in **Table 1**. The POS file contains data on characteristics of hospitals and other types of healthcare facilities, including the name and address of the facility and the type of Medicare services the facility provides, among other information. The data are collected through the CMS Regional Offices. General information about the POS Files is available at: https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/index.html

As described in more detail below in section 2b3. 4b., this performance measure does not adjust for social risk factors. However, we have conducted testing of social risk factors, and for this testing, we used data from the

Integrated Data Repository (IDR) file to capture patients' dual eligibility status. We extracted dual eligibility data from the IDR and added this variable to our primary dataset, the IRF-PAI:

- **Beneficiary Fact table (V2_MDCR_BENE_FCT) from the Integrated Data Repository (IDR):** CMS maintains the Integrated Data Repository (IDR), a high-volume data warehouse integrating Parts A, B, C, D, and DME claims, beneficiary and provider data sources, along with ancillary data such as contract information and risk scores.
- We used the IDR file to extract information on beneficiary dual eligibility status for social risk factor testing. These data are submitted by states to CMS and provide a monthly snapshot representing beneficiary characteristics as of set points in time. We used the BENE_DUAL_STUS_CD (Beneficiary Point of Sale Dual Status Code) that identifies the entitlement status for the dual eligible beneficiary. Missing data is rare and if it is missing for one month's data then the months before and after can be used. In this analysis, missing data for dual eligibility occurred for < 11 patient stays. General information about the IDR is available at: https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/IDR/.

**1.3. What are the dates of the data used in testing**?

For most testing reported in this document, we analyzed the records of patients discharged in calendar year 2017 (January 1, 2017 through December 31, 2017; 12 Months). For the Rasch analysis and internal consistency testing, we analyzed the records of patients discharged in fiscal year 2017 (October 1, 2016 through September 30, 2017; 12 Months).

**1.4. What levels of analysis were tested**? (*testing must be provided for <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

| **Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.20*)** | **Measure Tested at Level of:** |
|---|---|
| ☐ individual clinician | ☐ individual clinician |
| ☐ group/practice | ☐ group/practice |
| ☒ hospital/facility/agency | ☒ hospital/facility/agency |
| ☐ health plan | ☐ health plan |
| ☐ other: | ☐ other: |

**1.5. How many and which <u>measured entities</u> were included in the testing and analysis (by level of analysis and data source)**? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

**Inpatient Rehabilitation Facilities Included in the National IRF-PAI Data - Calendar Year 2017 Data**

Testing for this performance measure involved several types of data element, scale/instrument and computed performance score reliability and validity analyses, performance measure variability analyses, and social risk factor analysis. The unit of analysis for the data element and scale/instrument analyses is patient assessments or patient stays, and the unit of analysis for the computed performance measure score analyses is providers (i.e., IRFs). National data collection for the change in self-care functional status outcome measure began October 1, 2016 with the 2016 release (Version 1.4) of the IRF–Patient Assessment Instrument (IRF-PAI).

A total of 1,129 IRFs submitted IRF-PAI records during the testing period, January – December 2017. This represents 100% of this type of provider as defined by the Centers for Medicare and Medicaid Services.

**Table 1** displays the geographical location and facility characteristics of IRFs that reported IRF-PAI data for this performance measure. The majority of these IRFs are located in the southern United States (CMS Regions 4, 5, and 6) with over 20 percent in Region 6 (TX, LA, AR, OK, NM). The majority of IRFs are in urban settings (86.4%)

and under private ownership (56.7%). About 25 percent of IRFs are rehabilitation hospitals; most IRFs are units. Few IRFs are teaching facilities (12.1%). Facility size is presented based on the number of patient stays. Approximately 50 percent of facilities treated 296 or fewer patients who were discharged in 2017, and the range was one stay to 4,416 patient stays. Note that providers with less than 20 stays during the 12-month testing period are excluded from facility-level analyses presented below.

**Table 1. Number of IRFs Reporting by Facility Characteristics, Calendar Year 2017 (N=1,129)**

| Characteristic | Number (Percent) |
|---|---|
| CMS Region | |
| Region 1: CT, ME, MA, NH, RI, VT | 34 (3.0%) |
| Region 2: PR, VI, NY, NJ | 71 (6.3%) |
| Region 3: MD, DC, DE, WV, VA, PA | 122 (10.8%) |
| Region 4: NC, SC, TN, FL, GA, AL, KY, MS | 197 (17.5%) |
| Region 5: MI, MN, OH, IL, IN, WI | 209 (18.5%) |
| Region 6: TX, LA, AR, OK, NM | 233 (20.6%) |
| Region 7: MO, KS, IA, NE | 75 (6.6%) |
| Region 8: ND, UT, SD, WY, CO, MT | 43 (3.8%) |
| Region 9: NV, AZ, CA, HI, AS, Pacific Territories | 113 (10.0%) |
| Region 10: WA, AK, ID, OR | 32 (2.8%) |
| Urbanicity | |
| Rural | 154 (13.6%) |
| Urban | 975 (86.4%) |
| Ownership Type | |
| Government | 119 (10.5%) |
| Private | 640 (56.7%) |
| Non-profit | 370 (32.8%) |
| Rehabilitation hospital | 281 (24.9%) |
| Teaching Facility | 137 (12.1%) |
| Number of Patient Stays | |
| Decile 1: 1-104 | 125 (11.1%) |
| Decile 2: 105-152 | 114 (10.1%) |
| Decile 3: 153-192 | 113 (10.0%) |
| Decile 4: 193-240 | 108 (9.6%) |
| Decile 5: 241-296 | 112 (9.9%) |
| Decile 6: 297-361 | 112 (9.9%) |
| Decile 7: 362-480 | 112 (9.9%) |
| Decile 8: 481-694 | 111 (9.8%) |
| Decile 9: 695-1,022 | 111 (9.8%) |
| Decile 10: 1,024-4,416 | 111 (9.8%) |

Note: Values are reported as frequency (percent); Source: RTI analysis of IRF-PAI January – December 2017, and Provider of Service (POS) File 2017 (Program reference: LP57)

**Rasch Analysis Sample using National IRF-PAI Data – Fiscal Year 2017 Data**

As noted above, the reliability and validity testing that involved Rasch analysis and internal consistency testing was conducted using fiscal year 2017 data. This dataset included 1,126 IRFs. The characteristics of these IRFs are very similar to the provider data for the calendar year 2017 data reported above.

**Face Validity – Technical Expert Panel (TEP) Survey**

On March 27, 2017, RTI International, on behalf of the Centers for Medicare & Medicaid Services (CMS), convened an in-person Technical Expert Panel (TEP) in Baltimore, MD, to seek expert input on the Inpatient Rehabilitation Facilities Quality Reporting Program (IRF QRP) quality measures, including the functional status performance measures. A pre-TEP survey completed by 7 of the 10 TEP members provided us with some data to address face validity of the Change in Self-Care performance measure. The entities that the 10 TEP members represented were: 30% non-profit organization, 40% for-profit corporations, 20% government entities, and 10% professional association. Four of the TEP members have academic affiliations. The TEP members reported their residence in the following states: Alabama, California, Kentucky, Massachusetts, Minnesota, New York, North Carolina, Ohio, Texas.

**1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)**? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

**Total Number of Patients Included in the National IRF-PAI Data - Calendar Year 2017 Data**

IRFs submitted a total of 493,209 patient records for Medicare Part A and Medicare Advantage patient stays that ended during the testing time period (January 1 through December 31, 2017). The sociodemographic and stay-level characteristics of these Medicare patients are summarized in **Table 2**.

Patients older than the age of 65 accounted for nearly 87 percent of IRF patients. Female patients comprised just over half of the patients, nearly 80 percent of patients were white, and just under half were married. Overall, most patients lived with family or relatives prior to their IRF stay (65.4%) and more than 90 percent were admitted to the IRF from short-term general acute care hospitals. Stroke was the largest primary diagnosis group (23.3%) with debility and cardiorespiratory conditions (17.4%), fractures and other multiple trauma (11.7%), and other neurological conditions other than progressive neurological conditions (11.4%) as other major primary conditions. The majority of IRF patient stays ended with the patient discharged to home with or without care from a home health service organization (73.9%). About 15 percent of patients were discharged to other post-acute care settings, and 10 percent were discharged to a short-term general acute care hospital.

**Table 2. IRF Medicare Patient and Stay Characteristics, Patients Discharged in Calendar Year 2017 (N=493,209)**

| Characteristic | Number (Percent) |
|---|---|
| Age | |
| 64 and younger | 66,395 (13.5%) |
| 65 to 74 | 169,773 (34.4%) |
| 75 to 84 | 161,473 (32.7%) |
| 85 and older | 95,568 (19.4%) |
| Gender | |
| Male | 231,751 (47.0%) |
| Female | 261,458 (53.0%) |
| Race/Ethnicity* | |
| White | 390,837 (79.2%) |
| Black or African American | 54,971 (11.2%) |
| Hispanic or Latino | 23,361 (4.7%) |
| Asian | 7,876 (1.6%) |
| American Indian/Alaskan Native | 1,724 (0.4%) |
| Native Hawaiian/Pacific Islander | 1,954 (0.4%) |

| Characteristic | Number (Percent) |
|---|---|
| Marital Status | |
| Married | 231,146 (46.9%) |
| Widowed | 131,663 (26.7%) |
| Other** | 130,400 (26.4%) |
| Pre-Hospital Living With | |
| Living Alone | 1443,592 (29.1%) |
| Family/Relatives | 322,605 (65.4%) |
| Other*** | 27,012 (5.5%) |
| Primary Diagnosis | |
| Stroke | 114,722 (23.3%) |
| Hip or knee replacement | 20,882 (4.2%) |
| Non-traumatic brain dysfunction | 36,147 (7.3%) |
| Traumatic brain dysfunction | 20,912 (4.2%) |
| Non-traumatic spinal cord dysfunction | 21,516 (4.4%) |
| Traumatic spinal cord dysfunction | 4,570 (0.9%) |
| Progressive neurological conditions | 13,081 (2.7%) |
| Other neurological conditions | 56,170 (11.4%) |
| Fractures and other multiple trauma | 57,879 (11.7%) |
| Amputation | 14,622 (3.0%) |
| Other orthopedic conditions | 39,177 (7.9%) |
| Debility, cardiorespiratory conditions | 85,808 (17.4%) |
| Medically complex conditions | 7,494 (1.5%) |
| Admitted from Location | |
| Short-term General Hospital | 458,871 (93.0%) |
| Home (with or without home care) | 19,378 (3.9%) |
| Post-Acute Care**** | 11,517 (2.3%) |
| Other† | 2,937 (0.6%) |
| Not Listed | 506 (0.1%) |
| Discharge to Location | |
| Short-Term General Hospital | 49,206 (10.0%) |
| Home (with or without home care) | 364,486 (73.9%) |
| Post-Acute Care**** | 74,379 (15.1%) |
| Other† | 4,038 (0.8%) |
| Not Listed | 1,100 (0.2%) |

Note: Values are reported as frequency (percent)

*Percentages can add up to more than 100%; if more than 1 category was selected the patient is assigned to both categories.

**Includes divorced, separated, never married, and not assessed/no information.

***Includes friend, attendant, other person, and not assessed/no information.

**** Includes institutional settings: skilled nursing facilities, long-term care hospitals, and another IRF.

† Includes nursing homes, swing beds, critical access hospitals, hospice, inpatient psychiatric facilities, and other intermediate care settings.

Source: RTI analysis of IRF-PAI, January – December 2017 (Program reference: LP57).

**Rasch Analysis Sample using National IRF-PAI Data – Fiscal Year 2017 Data**

As noted above, the reliability and validity testing that involved Rasch analysis and internal consistency testing was conducted using fiscal year 2017 data. IRF-PAI data for 163,344 randomly selected IRF patients discharged in fiscal year 2017 were analyzed for the fit assessment and internal consistency. More than half of the IRF patients were female (53.3%) and 52.3% were 75 years old or older. Most were white (79.3%) and admitted to the IRF directly from an acute care hospital (93.2%).

**1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below**.

Most testing was conducted using national IRF-PAI data submitted by IRFs for all Medicare Part A and Medicare Advantage patients discharged in calendar year 2017 (**Tables 1** and **2**).

For the Rasch analyses and internal consistency analyses, we used a random subsample of the national data (n = 163,344) for patients discharged in fiscal year 2017. The Rasch analysis and internal consistency work include:

- Scale Construct Validity Testing - Item Difficulty Ordering
- Scale Validity Testing - Fit Assessment and Internal Consistency
- Item Validity Testing - Response Option Assessment

**1.8 What were the social risk factors that were available and analyzed**? For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

We examined whether 5 social risk factors affected the computed performance measure scores: 1) dual eligibility (patient-level variable); 2) race/ethnicity (patient-level variable); 3) living alone (patient-level variable); 4) urbanicity based on the patient's residence (community-level variable); and 5) socioeconomic status (SES) (community-level variable).

We selected the patient-level social risk factors based on our review of the literature showing functional outcomes can vary by race/ethnicity and by living situation. The selected community-level factors have been examined for other measures, but have been not addressed in the functional outcomes literature and thus the possible role and these factors have been unclear.

Dual eligibility data were derived from the Integrated Data Repository (IDR). We obtained race/ethnicity and living alone status from the IRF-PAI. Urbanicity was defined by cross-walking beneficiary residence ZIP codes (from the IRF-PAI) to Federal Information Processing Standard Publication (FIPS) codes,[1] then cross-walking FIPS codes to Rural-Urban Commuting Area Codes (RUCA_2013).[2] Socioeconomic status was determined using the Agency of Healthcare Research and Quality's SES Index[3] calculated based on the patient's residence ZIP Code Tabulation Area (ZCTA). ZCTA was found by cross-walking the beneficiary residence ZIP code with ZCTA. We used data from the 2016 American Community Survey (5-year file) to calculate AHRQ SES Index, with higher values indicating higher SES.

[1] https://www.huduser.gov/portal/datasets/usps_crosswalk.html

[2] https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes.aspx

**2a2. RELIABILITY TESTING**

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

**2a2.1. What level of reliability testing was conducted**? (*may be one or both levels*)

☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

**2a2.2. For each level checked above, describe the method of reliability testing and what it tests** (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

We report testing results throughout this document for data elements, the self-care scale/instrument and the computed performance measure score. To assist the reader in understanding the testing analysis and results, we begin by providing a brief overview of these components of the performance measure:

1. **Self-Care Data Elements:**
    a. There are 7 self-care data elements, which are included in IRF-PAI Section GG. Depending on the context, we sometimes refer to these data elements as "items" or "activities."
    b. The self-care data are collected at the time of admission and discharge using a 6-level rating scale (01 to 06), or activity not attempted codes if, for example, the activity was not attempted due to medical or safety concerns.
    c. Higher scores indicate higher ability (i.e., more independence)
    d. For the performance measure calculation, data element activity not attempted codes and missing data are recoded to 01.
2. **Admission and Discharge Self-Care Scores (Scale/Instrument)**
    a. An admission self-care scale score is created by summing the 7 self-care data element scores, after re-coding. The admission self-care score can range from 7 to 42 self-care units.
    b. A discharge self-care scale score is created by summing the 7 data element scores, after re-coding. The range of the discharge self-care score is 7 to 42 self-care units.
    c. For the Admission and Discharge Self-Care Scores, a score of 7 indicates the patient is dependent on a helper to perform all 7 self-care activities (i.e., data elements) and a score of 42 means the patient is independent on all 7 activities.
3. **Observed Change in Self-Care**
    a. An observed change in Self-Care score is calculated by subtracting the observed (unadjusted) Discharge Self-Care Score from the observed (unadjusted) Admission Self-Care Score.
    b. The potential range of the Observed Change in Self-Care Scores is -35 to + 35. Most patients are expected to have improved self-care abilities, and thus we observe mostly positive values.
4. **Calculated Performance Measure Score: Risk-Adjusted Change in Self-Care Score**
    a. The calculated performance measure score is a risk-adjusted Change in Self-Care Score. The risk-adjustment procedure is described in S.14. Calculation Algorithm/Measure Logic on the NQF Intent to Submit form and the attached file "IRF_Detailed_Function_QM_Specifications_2633_01-07-2019.docx".
    b. This performance measure estimates the risk-adjusted mean change in self-care score between admission and discharge for IRF patients. This performance measure does not have a simple form for the numerator and denominator.

**Computed Performance Measure Score Reliability – Split-half Reliability (unit of analysis is providers):** Split-half reliability was used to examine the reliability of the computed performance measure scores. The computed performance measure scores are the risk-adjusted change in self-care scores. For IRFs with fewer than 20 patient stays, computed performance measures are not displayed to the public, therefore, we included facilities with 20 or more stays in this analysis. We conducted split-half reliability by randomly

splitting each provider's patient stays into two groups and calculating correlations between the computed performance measure scores of the randomly divided groups. When a provider's data, after being randomly divided into two groups, show similar scores to one another, the performance measure score is more likely to reflect systematic differences in IRF provider quality rather than random variation. The Pearson Product-Moment Correlation (*r*), Spearman Rank Correlation (*ρ)*, and Intraclass Correlation Coefficient (ICC) were used to measure internal reliability. Intraclass correlations were also calculated by facility volume quartile to examine whether there were differences in performance measure reliability by IRF size.

**Self-Care Scale/Instrument Analysis - Internal Consistency (unit of analysis is patient assessments):** In addition to the provider-level reliability testing of the computed performance measure scores described above, we examined the internal consistency of the self-care scale/instrument scores for each patient-stay. Internal consistency provides a general assessment of how well the self-care data elements interrelate within the self-care scale/instrument. This internal consistency analysis is an indicator of the reliability of the self-care scale/instrument and is thus a test of the reliability of the data elements.

Internal consistency was assessed using the Cronbach's alpha coefficient, which is the average correlation of all possible half-scale divisions. Cronbach's alpha is a statistic frequently calculated when testing instrument or scale psychometrics. The Cronbach's alpha reliability estimate ranges from zero to one, with an estimate of zero indicating that there is no consistency of measurement among the items, and one indicating perfect consistency. Many cutoff criteria exist to determine whether or not a scale shows good consistency or whether the items "hang together" well. Nunnally (1978) indicated that Cronbach's alpha should be at least 0.90 for item sets used in decision making. The internal consistency from the Rasch analysis assesses items using the KR20 (a special case of Cronbach's alpha) estimate, with the same cut-off requirements

**Citation:** Nunnally, J. (1978). *Psychometric methods*. New York, NY: McGraw-Hill.

**Critical Data Elements Testing using CARE Tool Data (2014) – Inter-Rater Reliability, Video (Standardized Patient) Reliability and Validity Testing (unit of analysis is patients)**: In our 2014 NQF testing document, we described several types of data element and scale/instrument reliability and validity analysis using data collected by providers as part of the Post-Acute Care Payment Reform Demonstration (2007-2012). This reliability and validity testing included the self-care and mobility data elements, as well as data elements that are used as risk adjustors for this performance measure. For more information about the development and testing of the data elements and scale/instrument, please see:

- Gage BJ, Constantine R, Aggarwal MM, Bernard S, Munevar D, Garrity M, Deutsch A, et al. (June, 2012). *The Development of the Continuity Assessment Record and Evaluation (CARE) Tool: Final Report*. Prepared for the Centers for Medicare & Medicaid Services. Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/The-Development-and-Testing-of-the-Continuity-Assessment-Record-and-Evaluation-CARE-Item-Set-Final-Report-on-the-Development-of-the-CARE-Item-Set-Volume-1-of-3.pdf
- Gage BJ, Smith LM, Ross J, Coots LA, Shamsuddin KM, Deutsch A, Mallinson T, Reilly KE, Abbate JH, Gage-Croll Z. (August, 2012). *The development and testing of the Continuity Assessment Record and Evaluation (CARE) Item Set: Final Report on Reliability Testing, Volume 2 of 3*. Prepared for Centers for Medicare & Medicaid Services. Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/The-Development-and-Testing-of-the-Continuity-Assessment-Record-and-Evaluation-CARE-Item-Set-Final-Report-on-Reliability-Testing-Volume-2-of-3.pdf
- Gage BJ, Deutsch A, Smith LM, Schwartz C, Ross J, Coots LA, Reilly KE, Abbate JH, Shamsuddin KM, Silver BC, et al. (September, 2012). *The Development and Testing of the Continuity Assessment Record and Evaluation (CARE) Item Set: Final Report on CARE Item Set and Current Assessment Comparisons, Volume 3 of 3*. Prepared for Centers for Medicare & Medicaid Services. Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-

Quality-Initiatives/Downloads/The-Development-and-Testing-of-the-Continuity-Assessment-Record-and-Evaluation-CARE-Item-Set-Final-Report-on-the-Development-of-the-CARE-Item-Set-and-Current-Assessment-Comparisons-Volume-3-of-3.pdf

- Smith LM, Deutsch A, Hand LB, Etlinger AL, Ross J, Abbate JH, Gage-Croll Z, Barch D, Gage BJ. (September, 2012). *Continuity Assessment Record and Evaluation (CARE) Item Set: Additional Provider-Type Specific Interrater Reliability Analyses*. Prepared for Centers for Medicare & Medicaid Services. Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/Continuity-Assessment-Record-and-Evaluation-CARE-Item-Set-Additional-Provider-Type-Specific-Interrater-Reliability-Analyses.pdf
- Smith LM, Deutsch A, Barch D, Ross J, Shamsuddin KM, Abbate JH, Schwartz C, Gage BJ. (September, 2012). *Continuity Assessment Record and Evaluation (CARE) Item Set: Video Reliability Testing*. Prepared for Centers for Medicare & Medicaid Services. Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/Continuity-Assessment-Record-and-Evaluation-CARE-Item-Set-Video-Reliability-Testing.pdf
- Gage BJ, Morley MA, Smith LM, Ingber MJ, Deutsch A, Kline TL, Dever JA, Abbate JH, Miller RD, Lyda-McDonald B, Kelleher CA, Garfinkel DB, Manning JR, Murtaugh CM, Stineman MG, Mallinson T. (March, 2012). *Post-Acute Care Payment Reform Demonstration: Final Report Volumes 1-4*. Prepared for the Centers for Medicare and Medicaid Services. Available at: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Reports/Research-Reports-Items/PAC_Payment_Reform_Demo_Final.html

For more information on the history of the development of this functional status performance measure, please visit CMS's Post-Acute Care Quality Initiatives Function Measures website: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Functional-Measures-.html

**Summary of critical data element reliability testing:**

The inter-rater reliability of the data elements was tested in a subset of 34 providers (acute hospitals, HHAs, IRFs, LTCHs, and SNFs) distributed across 11 geographic areas. Each provider completed a duplicate admission or discharge assessment on 10–20 patients. The overall sample size was 450 for self-care items. Weighted kappa values for the self-care items range between 0.798 for eating to 0.869 for upper-body dressing. Unweighted kappas ranged from 0.598 for oral hygiene to 0.634 for upper-body dressing. Unweighted overall kappas ranged from 0.636 (toileting) to 0.598 (oral hygiene). In summary, kappa statistics indicated substantial agreement of data element codes among raters.

For the video reliability study, clinicians assessed "standardized" patients presented through a videotape of a patient assessment. This ensured that the same information was presented to each clinician and allowed examination of scoring among different clinicians examining the "same" patient. The video reliability study indicated substantial agreement with the mode and clinical team among all items, typically upwards of 70%. The notable exception to this trend exists among the clinicians in the "Other" category (mostly LPNs); they consistently had the lowest levels of agreement among self-care items, ranging from 50 to 72%. For the toileting and dressing items, the agreement with the clinical team was lower than with the mode. This occurred because the clinical team response differed from the mode for these three items in either one or two videos. Because the clinical team response and mode were identical on most of the videos, agreement was still quite high for these items.

Please see Appendix B for the additional details about the inter-rater reliability and video reliability testing.

**2a2.3. For each level of testing checked above, what were the statistical results from reliability testing**? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

**Computed Performance Measure Score Reliability (unit of analysis is provider):** Split-half analysis results indicated strong, positive correlations (r = 0.903, ρ = 0.884, ICC= 0.903, p < 0.0001) between the IRF providers' randomly divided groups' computed performance measure scores for the Change in Self-Care performance measure, providing strong evidence of measure reliability. As shown in **Table 3**, ICCs remained strong when stratifying by provider volume quartile, with ICCs for the volume quartiles ranging from 0.836 (20-174 discharges) to 0.965 (568 - 4,416 discharges).

**Table 3. Interclass Correlation Coefficient by IRF Volume, Calendar Year 2017 (N=1,117)**

| Volume Quartile | Number of IRFs | ICC |
|---|---|---|
| **Quartile 1: 20 - 174** | 280 | 0.836 |
| **Quartile 2: 175 - 295** | 278 | 0.899 |
| **Quartile 3: 296 - 566** | 280 | 0.944 |
| **Quartile 4: 568 - 4,416** | 279 | 0.965 |
| **Total** | 1,117 | 0.903 |

Note: Providers with < 20 stays during the 12-month testing period are excluded.
Source: RTI analysis of IRF-PAI January – December 2017 (Program reference: MV52).

**Scale/Instrument Reliability - Internal Consistency (unit of analysis is patient stays):** Analysis of the self-care data showed good reliability statistics. The overall Cronbach's alpha is 0.94.

**2a2.4 What is your interpretation of the results in terms of demonstrating reliability**? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The analysis of calendar year 2017 data show that provider-level reliability of the computed performance measure scores was strong overall and when stratified by provider volume. The patient-level analysis of fiscal year 2017 data of the scale/instrument reliability showed very good reliability.

Critical data element inter-rater reliability and video reliability testing found substantial reliability overall.

---

**2b1. VALIDITY TESTING**

**2b1.1. What level of validity testing was conducted**? (*may be one or both levels*)

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

**2b1.2. For each level of testing checked above, describe the method of validity testing and what it tests** (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

**Scale/Instrument Content Validity - Similarity of Data Elements Across Other Self-Care Assessment Instruments:** Patient functioning is a construct that is often measured based on patient abilities, and the activities (data elements) included in functional assessment instruments vary. We compared the list of Section

GG data elements used to calculate the Change in Self-Care performance measure with self-care data elements included on other functional assessment instruments.

**Face Validity – Technical Expert Survey** - On March 27, 2017, RTI International, on behalf of the Centers for Medicare & Medicaid Services (CMS), convened an in-person Technical Expert Panel (TEP) in Baltimore, MD, to seek expert input on the Inpatient Rehabilitation Facilities Quality Reporting Program (IRF QRP) quality measures, including the functional status performance measures. Prior to the TEP meeting, TEP members provided feedback on the importance, scientific soundness and usability of each of the performance measures using a 5-level Likert scale (high, moderately high, neutral, moderately low, low).

**Data Element Construct Validity – Observed Discharge Self-Care Scores and Discharge Destination (unit of analysis is patient stays):** We tested the validity of the self-care data by examining the discharge function scores and whether patients were discharged to a community destination. IRFs provide intensive rehabilitation services to patients with a goal of maximizing patient functioning so that the patient can be ideally discharged home and avoid institutionalization. IRF patients who have higher abilities are more likely to be discharged to their home or another community-based setting compared to patients discharged to another post-acute care setting (e.g., skilled nursing facility, long-term care hospital), nursing home, hospice, or an acute-care hospital. Therefore, we tested the construct validity of the self-care data by examining the relation between discharge self-care scores and being discharged to the community, after excluding incomplete stays.

**Scale/Instrument Construct Validity – Observed Discharge Self-Care Scores and Discharge Destination (unit of analysis is patient stays):** We tested the validity of the scale/instrument scores by examining the discharge self-care scale scores and whether patients were discharged to a community destination. We ran a logistic regression model to examine the association between discharge self-care scores and the odds of a community discharge.

**Scale/Instrument Construct Validity - Data Element (Item) Difficulty Ordering Using Rasch Analysis (unit of analysis is patient assessment data):** Rasch analysis uses item data to determine how well items in a scale/instrument function together to measure a construct. In its base form, the Rasch model assumes that the probability of a code for a given item is a function of the patient's self-care ability and the item's difficulty (how hard the activity is to accomplish independently). The Rasch extension that accounts for multiple response options also considers the difficulty of moving from one code category to another (i.e., a threshold). The information resulting from this function is interval in nature and expressed on the log-odds scale. Also, as part of the analysis, Rasch methodology places persons and the items of interest on a "ruler" to enable evaluations of how well the items work together, how difficult each item is relative to the other items in the scale/instrument, and how items are ordered from easy to difficult. We used Rasch measurement analysis to examine the self-care items. We report IRF analysis results using a Rasch-derived self-care ruler that was developed using data from IRFs, skilled nursing facilities and long-term care hospitals. Using the Rasch-derived cross-setting "ruler" allows comparability of self-care item functioning within and across settings.

The ordering of items from easy (bottom) to difficult (top) provides the analysis-established item difficulty hierarchy. This hierarchy can be evaluated against item design specifications (i.e., the intended construction of the items to be easy or difficult) and against expert clinical opinions as an indication of construct validity. If items are positioned into unexpected locations on the hierarchy, then the content of the items should be evaluated further and potentially modified.

**Data Element (Item) and Scale/Instrument Validity - Fit Assessment Analysis (unit of analysis is patient assessment data):** Rasch analysis produces fit statistics that reflect whether unexpected responses are being coded for items within the scale/instrument. The Rasch model expects the difficult items to be harder (that is, have a greater need for assistance) for all patients. In a similar way, patients with higher functional abilities are generally expected to need less assistance on all items. Items that don't seem to function this way could show misfit, reflecting unexpected responses. There are two categories of fit, one designed more for outliers (outfit) and one designed for response unexpectedness near the item's difficulty (infit). In general, a cut-off appropriate for statistically determining item misfit is infit and outfit mean square values are above 1.4 when

looking at multiple-point response scales. Items with fit values above 1.4 are unproductive for measurement but are not unusually "noisy" or degrade measurement. Mean square values greater than 2.0 may potentially degrade measurement (Wright and Linacre, 1994). Misfit seen near the item difficulty, or large values of infit, are concerning because they indicate noise (unexpected responses) where the item should be the most productive for measurement.

**Data Element (Item) and Scale/Instrument Validity - Response Option Assessment Using Rasch Analysis (unit of analysis is patient assessment data):** Rasch analysis output reports the number and percent of patients by score level (06 - Independent to 01 - Dependent) for each item and the average self-care ability (i.e., scale-level ability) of those patients. This allows us to examine if the 6-point rating scale is operating as intended for the self-care items. In general, we expect that patients who have lower ability overall would have lower ability levels (i.e., lower scores) for each item. Therefore, the average self-care ability calibration (scale-level ability measure reported in logits) associated with the more dependent scores would be lower than those associated with the more independent scores.

**Citation:**

Wright BD, Linacre JM (1994) Reasonable mean-square fit values. *Rasch Measurement Transactions.* 8:3 p.370. http://www.rasch.org/rmt/rmt83b.htm

**Computed Performance Measure Score Validity – Association with The Joint Commission Stroke Rehabilitation Certification Status (unit of analysis is providers):** The goal of measuring performance is to make valid (credible) conclusions about quality (NQF Committee Guidebook). To examine the validity of the Change in Self-Care computed performance measure score, we conducted analyses using a structural measure of quality, whether or not an IRF obtained The Joint Commission's Disease Specific Certification for Stroke Rehabilitation. As previously noted in Table 1, stroke is the most common primary medical condition for patients admitted to IRFs, therefore stroke patient outcomes influence IRF performance measure scores. The Joint Commission's Disease-Specific Care Certification evaluates clinical programs addressing: 1) Compliance with consensus-based national standards; 2) Effective use of evidence-based clinical practice guidelines to manage and optimize care; and 3) An organized approach to performance measurement and improvement activities. According to The Joint Commission, an entity that achieves Disease-Specific Certification has thoroughly demonstrated a high level of care for patients with that condition. We downloaded data from The Joint Commission's website and we used an 'effective date' to identify IRFs that were certified during the calendar year 2017. More information about disease-specific certification, please see: https://www.jointcommission.org/certification/dsc_physical_medicine_rehabilitation.aspx

Our first analysis compared the mean and median computed performance measure scores for IRFs with and without stroke rehabilitation disease-specific certification using a t-test and Kruskal-Wallis H test. We expected that IRFs with certification would achieve higher mean and median performance measure scores compared to IRFs without certification. Second, we divided the IRF data into quintiles based on the performance measure scores and calculated the percentage of IRFs with certification by quintile. We expected that IRFs with the best performance scores (quintile 5) would have a higher percentage of certified IRFs compared to the IRFs in quintile 1 with the least favorable performance measure scores.

**2b1.3. What were the statistical results from validity testing**? (*e.g., correlation; t-test*)

**Content Validity: Similarity of Data Elements Across Other Self-Care Assessment Instruments.**

Many functional status scales have been developed for research and clinical use. To address content validity, we have provided a table listing activities (data elements) used to calculate the Change in Self-Care performance measure and data elements included in other functional assessment scales. **Table 4**, shows that the Section GG self-care activities cover a wide range of self-care activities and that many of the activities included on other instruments (e.g., Eating, Toileting hygiene) are included in Section GG.

**Table 4. Comparison of Selected Self-Care Activities (Data Elements) for the Change in Self-Care Performance Measure and Other Functional Assessment Instruments.**

| Activity (Data Elements) | Change in Self-Care Measure: Section GG Self-Care Items | Barthel Index | FIM® Instrument | Katz ADL Scale | FAM with FIM® Instrument | RICFAS with FIM® Instrument | Minimum Data Set: Section G | Upper Extremity Functional Index |
|---|---|---|---|---|---|---|---|---|
| **Eating**<br>**ICF =** Drink d560<br>Eat d550 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |
| **Grooming/personal or oral hygiene**<br>**ICF =** Caring for teeth d5201 | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ |
| **Toileting hygiene**<br>**ICF =** Regulating Urination d5300<br>Regulating Defecation d5301 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |
| **Bathing/showering**<br>**ICF = d**5100 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |
| **Dressing: lower body**<br>**ICF =** Putting on clothes d5400<br>Taking off clothes d5401 | ✓ | | ✓ | | ✓ | ✓ | | ✓ |
| **Dressing: upper body**<br>**ICF =** Putting on clothes d5400<br>Taking off clothes d5401 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Dressing: footwear**<br>**ICF =** Putting on clothes d5400<br>Taking off clothes d5401 | ✓ | | | | | | | ✓ |

Notes: ADL = activity of daily living; ICF = International Classification of Functioning; FAM = Functional Assessment Measure; RICFAS = Rehabilitation Institute of Chicago Functional Assessment Scale

**Face Validity – Technical Expert Survey:** For the Change in Self-Care performance measure, 71% of TEP members rated the Measure Importance as High or Moderately High; 57% rated the Scientific Soundness as High or Moderately High, and 57% Rated Usability of the Measure as High. We note that this survey was conducted prior to the TEP meeting, and thus represents perceptions before the TEP discussions about the measure details and measure testing results. In addition, this TEP occurred approximately 8 months after the implementation of data collection when confidential feedback reports were not yet available to providers. Finally, with the goal of learning from experts in order to drive measure improvement efforts, for this TEP we invited representatives from organizations that had previously given feedback on the measures and that had competing measures. Thus, full support for the measure was not an expected outcome of the pre-TEP survey, and the survey provided TEP members an opportunity to give constructive feedback based on their initial perceptions before participating in the panel.

**Data Element Construct Validity: Observed Discharge Functional Ability and Discharge Destination (unit of analysis is patient stays):** As shown in **Table 5,** patients with higher discharge scores (from 01 – Dependent to 06 – Independent), are more likely to be discharged to the community, as expected. This occurs for each self-care data element for all score levels, with the exception of the data element Eating level 1, which has a slightly higher percentage compared to level 2. Also expected, for each of the self-care data elements (**Table 5**), patients who were coded as 06 - Independent, a high percentage were discharged to the community (89.6% for Eating to 97.3% for Shower/Bathe Self).

Findings and Interpretation: Self-care data were positively associated with discharge destination, as expected. Specifically, we found patients who had higher observed scores at discharge were more likely to be discharged to a community setting, which supports the validity of the item data measuring functional abilities in the IRF population.

**Table 5. Observed Discharge Self-Care Data Element Scores and Discharge Location (n=437,619)**

| Data Element/Score | Discharged to Community | |
|---|---|---|
| **GG0130A3: Self-Care - Eating** | | |
| 01-Dependent | 2,131 (43.8%) | |
| 02-Substantial/maximal assistance | 1,114 (40.7%) | |
| 03-Partial/moderate assistance | 2,294 (48.7%) | |
| 04-Supervision or touching assistance | 15,109 (61.5%) | |
| 05-Setup or clean-up assistance | 55,914 (68.0%) | |
| 06-Independent | 281,212 (89.6%) | |
| **GG0130B3: Self-Care - Oral Hygiene** | | |
| 01-Dependent | 1,227 (34.6%) | |
| 02-Substantial/maximal assistance | 1,508 (38.8%) | |
| 03-Partial/moderate assistance | 4,440 (47.7%) | |
| 04-Supervision or touching assistance | 28,160 (66.3%) | |
| 05-Setup or clean-up assistance | 62,451 (68.5%) | |
| 06-Independent | 260,236 (91.8%) | |
| **GG0130C3: Self-Care - Toileting Hygiene** | | |
| 01-Dependent | 11,071 (38.7%) | |
| 02-Substantial/maximal assistance | 14,430 (52.8%) | |
| 03-Partial/moderate assistance | 25,697 (67.2%) | |
| 04-Supervision or touching assistance | 72,684 (79.8%) | |
| 05-Setup or clean-up assistance | 31,936 (84.0%) | |
| 06-Independent | 202,239 (96.2%) | |
| **GG0130E3: Self-Care - Shower/Bathe Self** | | |
| 01-Dependent | 3,155 (34.5%) | |
| 02-Substantial/maximal assistance | 8,215 (44.1%) | |
| 03-Partial/moderate assistance | 45,471 (63.6%) | |
| 04-Supervision or touching assistance | 97,907 (82.5%) | |
| 05-Setup or clean-up assistance | 55,335 (88.3%) | |
| 06-Independent | 145,718 (97.3%) | |

| Data Element/Score | Discharged to Community | |
|---|---|---|
| **GG0130F3: Self-Care - Upper Body Dressing** | | |
| 01-Dependent | 2,124 (35.5%) | |
| 02-Substantial/maximal assistance | 5,792 (43.4%) | |
| 03-Partial/moderate assistance | 24,826 (60.2%) | |
| 04-Supervision or touching assistance | 43,858 (73.2%) | |
| 05-Setup or clean-up assistance | 67,130 (76.4%) | |
| 06-Independent | 215,723 (94.9%) | |
| **GG0130G3: Self-Care - Lower Body Dressing** | | |
| 01-Dependent | 8,562 (38.5%) | |
| 02-Substantial/maximal assistance | 17,755 (52.5%) | |
| 03-Partial/moderate assistance | 39,276 (69.1%) | |
| 04-Supervision or touching assistance | 75,410 (81.2%) | |
| 05-Setup or clean-up assistance | 41,270 (87.0%) | |
| 06-Independent | 177,127 (97.2%) | |
| **GG0130H3: Self-Care - Putting on/Taking Off Footwear** | | |
| 01-Dependent | 16,689 (46.5%) | |
| 02-Substantial/maximal assistance | 18,191 (57.3%) | |
| 03-Partial/moderate assistance | 34,789 (72.6%) | |
| 04-Supervision or touching assistance | 57,825 (80.7%) | |
| 05-Setup or clean-up assistance | 48,402 (84.3%) | |
| 06-Independent | 180,806 (96.6%) | |

Notes: Values reported as frequency (percent); Incomplete stays are excluded; Activity not attempted codes not shown.
Source: RTI analysis of IRF-PAI, January– December 2017. (Program reference: LP63).

**Scale/Instrument Construct Validity: Observed Discharge Functional Ability and Discharge Destination (unit of analysis is patient stays). Table 6** displays the single variable logistic regression results with observed discharge self-care scale scores as the independent variable and a dichotomous dependent variable indicating whether the IRF patient was discharged to the community or not. The self-care scale score is the sum of the 7 self-care data element scores after recoding; the discharge self-care scale scores can range from 7 to 42. The results show that, on average, a one-unit increase in discharge self-care score is associated with a 16 percent increase in the odds of being discharged to the community (OR = 1.159; *p*-value <0.001).

Findings and Interpretation: Self-care scale/instrument scores were positively associated with discharge destination, as expected. Specifically, we found patients who had higher observed scores at discharge were more likely to be discharged to a community setting, which supports the validity of the scale/instrument data measuring functional abilities in the IRF population.

**Table 6. Coefficient and Odds Ratio for Discharge to Community Model (n=437,619)**

| Independent Variable | Value | 95% Confidence Interval |
|---|---|---|
| Observed Discharge Self-Care Score | | |
| Coefficient | 0.148 | |
| Odds Ratio | 1.159 | 1.158 – 1.161 |

Note: Observed discharge self-care score range = 7 – 42; Incomplete stays were excluded.
Source: RTI analysis of IRF-PAI, January – December 2017. (Program reference: LP63).

**Scale/Instrument Construct Validity: Data Element (Item) Difficulty Ordering Using Rasch Analysis (unit of analysis is patient assessment data):** We used Rasch analysis to determine how well the self-care items work together to measure the construct of self-care. Rasch analysis creates a self-care ruler using log odd units (i.e., logits) centered at the value 0. A "logit" (a contraction of "Log-Odds Unit") is a linear scale We report IRF testing results using a Rasch-derived self-care ruler that was developed using data from IRFs, skilled nursing facilities and long-term care hospitals. The analysis of the Section GG self-care data show that the placement of each self-care item on the cross-setting self-care "ruler" make sense clinically and are consistent with previous analyses of other functional assessment scales. That is, the order of items from easy to difficult (item hierarchy), is consistent with task difficulties. The order of the items by difficulty level, with the hardest activity listed first, is as follows:

Footwear (most difficult activity)

Lower Body Dressing

Shower

Toileting Hygiene

Upper Body Dressing

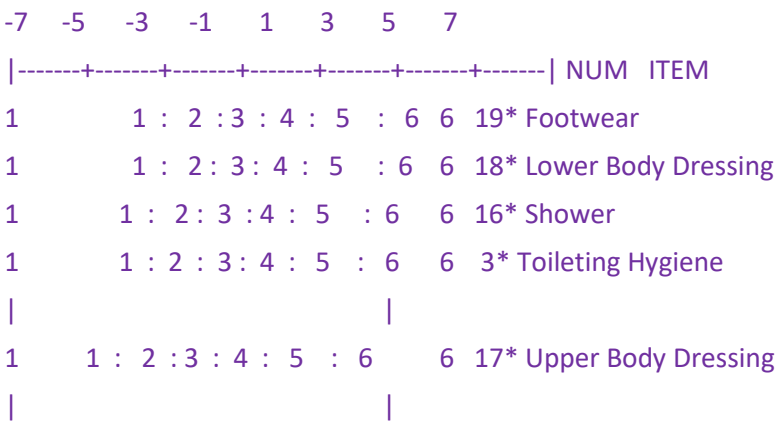
Oral Hygiene

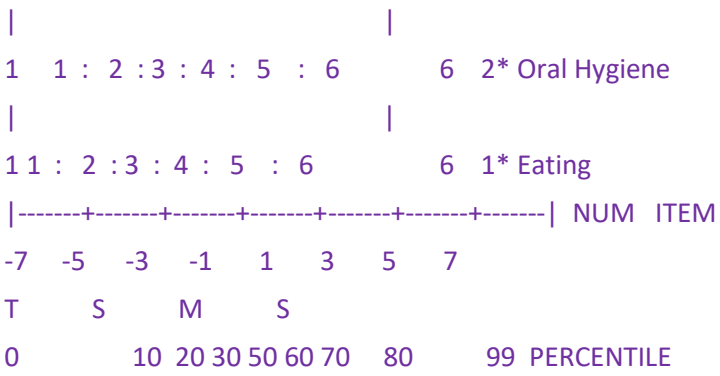
Eating (easiest activity)

**Figure 1** reports the item hierarchy, the patient distribution and the rating scale scores in one graphic. In addition, **Figure 1** is presented on the Rasch-derived self-care ruler, expressed in logits and centered at a value of 0, as described previously. It shows the overall expected score placement on the self-care "ruler" for each item. The ruler values, ranging from -7 to +7 logits, are shown on the top and bottom vertical lines. The difficulty order (item hierarchy), from easy (bottom) to difficult (top), is shown on the right side of the graphic. For each item presented on the right, the overall expected placement of the score options (from "1" for "dependent" to "6" for "independent") are shown along the ruler. Each item is presented on a row and the scores begin with the most dependent (represented by the "1") on the far-left graphic boundary and the most independent (represented by "6") on far-right graphic boundary. Finally, the threshold between two score options is represented by a colon (:) and is where a patient has an equal chance of being in either the higher or lower category. Use of the "ruler" allows visualization of the scores for each self-care item in relation to the scores of other self-care items. The letters at the bottom of **Figure 1** describe the distribution of people along the ruler, where "M" is the average of the sample and "S" and "T" are one and two times the standard deviation around that average, respectively. The percentile values represent the distribution of patients along the "ruler."

**Findings and Interpretation:** The item hierarchy listing and **Figure 1** illustrate that the self-care items fall along the cross-setting "ruler" as expected and are consistent with clinical findings from applications in the field and other functional assessment instruments.

**Figure 1. Self-Care IRF Items – Anchored on the Cross-Setting Self-Care Ruler**






**Scale/Instrument Validity - Fit Assessment Using Rasch Analysis (unit of analysis is patient assessment data):** Ideal measurement construction would mean data fit the Rasch model exactly. In reality, empirical data will differ from the model. Rasch fit statistics describe how well the observed data (e.g. patient's scores on the self-care items) fit the model, and characterize the magnitude that unexpected scores (i.e., unmodelled noise) are found in the data. Fit statistics have an expected value of 1.0 and can range from 0 to infinity. Values lower than 1.0 indicate overfit (over prediction) of the Rasch model and values greater than 1.0 indicate underfit of the model (e.g., noise). There are two categories of fit. Outfit is designed more for outliers (when a patient's unexpected code is for an item that is relatively easy or hard for that patient); Infit is designed for unexpected codes near the item's difficulty (when a patient's code is for an item is near that person's ability). Values greater than 2.0 may potentially degrade measurement (Wright and Linacre, 1994). Overall, the self-care items are coded as expected. **Table 7** reports fit statistics for the self-care items and shows that one item, Eating, had fit statistics above 2.00. The item Eating has been found to misfit in other functional assessment scales, and this misfit may be related to patients who have a swallowing impairment.

**Table 7. Fit Statistics for the Self-Care Items (n = 321,392)**

| | **IRF – Anchored (Cross-Setting Ruler)** | |
|---|---|---|
| Item | Infit mean square | Outfit mean square |
| GG0130A: Eating | 1.56 | 2.38 |
| GG0130B: Oral Hygiene | 1.14 | 1.16 |
| GG0130C: Toileting Hygiene | 1.15 | 1.15 |
| GG0130E: Shower/Bathe Self | 0.79 | 1.04 |
| GG01030F: Upper Body Dressing | 0.98 | 0.99 |
| GG0130G: Lower Body Dressing | 0.64 | 0.67 |
| GG0130H: Putting On/Taking Off Footwear | 1.19 | 1.14 |

**Data Element (Item) and Scale/Instrument Validity - Response Option Assessment Based on Rasch Analysis (unit of analysis is patient assessments):** Rasch analyses provide information on how many patients are coded in each score category (i.e., independent to dependent) for each item and the average ability (or skill level) of those individuals on the construct of interest. Evaluations of patient ability by score category indicate that rating scale use is as expected, with patients with higher item scores are, on average, higher ability patients. For our data, we anticipate that for each item, patients with higher scores (01 to 06) should have higher Rasch logit self-care values (Rasch self-care logit values range from -7 to +7). Likewise, it is expected that lower ability persons would generally be observed in the more dependent categories (substantial assistance, etc.). Therefore, the average ability (or skill level) estimate associated with the more dependent scores would be lower than ability estimates associated with the more independent scores. We combined admission and discharge data for each item in order to ensure a range of patient ability is represented in the analyses.

As shown in **Table 8**, for each item, patients who are coded with higher scores have higher overall self-care ability, as expected.

**Table 8. Distribution of Combined Admission and Discharge Scores and Average Ability Estimate by Response Code for Each Self-Care Item (n = 321,392)**

| Item | Score (Response Code)* Higher Score = Higher Ability | Number of Patients | Percent of Patients by Item | Average Self-Care Ability of Patients (- 7 to +7 Logit Scale) Higher Value = Higher Ability) |
|---|---|---|---|---|
| Eating | | | | |
| | 01 | 7503 | 2 | -4.33 |
| | 02 | 4778 | 2 | -3.21 |
| | 03 | 10118 | 3 | -2.09 |
| | 04 | 32476 | 10 | -0.83 |
| | 05 | 103346 | 33 | -0.03 |
| | 06 | 156927 | 50 | 3.16 |
| Oral hygiene | | | | |
| | 01 | 7490 | 2 | -5.04 |
| | 02 | 7832 | 3 | -2.84 |
| | 03 | 20976 | 7 | -1.61 |
| | 04 | 58049 | 19 | -0.30 |
| | 05 | 112306 | 36 | 0.30 |
| | 06 | 106575 | 34 | 4.44 |
| Toileting hygiene | | | | |
| | 01 | 50601 | 16 | -2.18 |
| | 02 | 44380 | 14 | -0.84 |
| | 03 | 49142 | 16 | 0.08 |
| | 04 | 68799 | 22 | 1.13 |
| | 05 | 21247 | 7 | 2.18 |
| | 06 | 75748 | 24 | 5.52 |
| Shower/bathe self | | | | |
| | 01 | 19958 | 7 | -3.24 |
| | 02 | 39856 | 13 | -1.42 |
| | 03 | 95554 | 32 | -0.09 |
| | 04 | 69272 | 23 | 1.60 |
| | 05 | 27569 | 9 | 3.22 |
| | 06 | 50838 | 17 | 6.47 |
| Upper body dressing | | | | |
| | 01 | 15196 | 5 | -3.81 |
| | 02 | 28103 | 9 | -1.82 |
| | 03 | 62352 | 20 | -0.63 |
| | 04 | 57758 | 18 | 0.40 |
| | 05 | 73505 | 23 | 1.18 |
| | 06 | 79272 | 25 | 5.46 |

| Item | Score (Response Code)* Higher Score = Higher Ability | Number of Patients | Percent of Patients by Item | Average Self-Care Ability of Patients (- 7 to +7 Logit Scale) Higher Value = Higher Ability) |
|---|---|---|---|---|
| Lower body dressing | | | | |
| | 01 | 53550 | 17 | -2.22 |
| | 02 | 63504 | 20 | -0.71 |
| | 03 | 59284 | 19 | 0.39 |
| | 04 | 56764 | 18 | 1.51 |
| | 05 | 21514 | 7 | 2.86 |
| | 06 | 61990 | 20 | 6.26 |
| Putting on/taking off footwear | | | | |
| | 01 | 78309 | 25 | -1.65 |
| | 02 | 51625 | 17 | -0.53 |
| | 03 | 42052 | 13 | 0.55 |
| | 04 | 44284 | 14 | 1.37 |
| | 05 | 31622 | 10 | 2.35 |
| | 06 | 64390 | 21 | 6.09 |

Note: Activity not attempted/did not occur codes are not included in this analysis.

*Response categories are defined as: 1 – Dependent; 2 – Substantial/maximal assistance; 3 - Partial/moderate assistance; 4 - Supervision or touching assistance; 5 - Setup or clean-up assistance; and 6 - Independent.

**Computed Performance Measure Score Validity – Association with The Joint Commission Stroke Rehabilitation Certification Status (unit of analysis is providers):** We compared the mean and median computed performance measure scores for IRFs with and without stroke rehabilitation disease-specific certification. We also divided the IRF data into quintiles based on the performance measure scores and calculated the percentage of IRFs with certification by quintile.

**Table 9** shows that IRFs with certification achieved slightly higher mean and median performance measure scores compared to IRFs without certification (mean: 12.1 and 11.4 and $p < .0001$; median 12.0 and 11.3 and $p < .001$).

**Table 9. Mean and Median Change in Self-Care Computed Performance Measure Score (CY 2017) by Stroke Rehabilitation Disease Specific Certification Status (2017) (n = 1,117)**

| Change in Self-Care Performance Measure Score | Stroke Rehabilitation Disease Specific Certification Status (2017) | | |
|---|---|---|---|
| | No (n=941) | Yes (n=176) | p-value* |
| Mean (SD) | 11.4 (1.7) | 12.1 (1.4) | < 0.001 |
| Median (IQR) | 11.3 (2.1) | 12.0 (1.6) | < 0.001 |

Note: SD=Standard deviation; IQR=Interquartile range; Providers with <20 stays during the 12-month testing period are excluded from facility-level analyses.
*T-test was run to determine statistically significant differences for the mean scores; The Kruskal-Wallis H test was run to determine statistically significant differences for the median scores.
Source: RTI analysis of IRF-PAI, January – December 2017. (Program references: AD01)

**Table 10** shows that the top 3 quintiles, which included the IRFs with the best performance scores, had the highest percentage of certified IRFs (29.0%, 25.0%, and 25.6%) compared to the lowest quintile with the lowest performance measure scores (9.1%).

**Table 10. Percent of IRF with Stroke Rehabilitation Disease Specific Certification by Computed Performance Measure Score (CY 2017) Quintiles (n = 1,117)**

| Quintile Group Based on Performance Measure Score: Best to Worst | Stroke Rehabilitation Disease Specific Certification Status (2017)* | |
|---|---|---|
| | No (n=941) | Yes (n=176) |
| Quintile 5: 12.9-17.5 (best performance scores) | 177 (18.8%) | 51 (29.0%) |
| Quintile 4: 11.9-12.8 | 167 (17.7%) | 44 (25.0%) |
| Quintile 3: 11.1-11.8 | 190 (20.2%) | 45 (25.6%) |
| Quintile 2: 10.2-11.0 | 200 (21.3%) | 20 (11.4%) |
| Quintile 1: 5.4-10.1 (worst performance scores) | 207 (22.0%) | 16 (9.1%) |

Note: Providers with <20 stays during the 12-month testing period are excluded from facility-level analyses.
*Chi square test results: $p < .0001$
Source: RTI analysis of IRF-PAI, January – December 2017. (Program references: AD01)

**2b1.4. What is your interpretation of the results in terms of demonstrating validity**? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The activities (data elements) included in the Section GG self-care scale/instrument cover a wide range of patient functioning and key activities included in many other functional assessment instruments, supporting content validity of the scale.

Prior to their participation in the TEP, the panel members were surveyed on their initial perceptions of the Change in Self-Care performance measure. Most experts convened indicated the performance measure was important, scientifically sound, and able to be used by providers, patients, and the general public.

We found that patients who had higher observed discharge scores for the self-care data elements were more likely to be discharged to the community, as expected. Results also showed that the observed self-care scale/instrument scores were significantly associated with being discharged to the community.

The difficulty order of the self-care data elements makes sense clinically and are consistent with previous analyses of the self-care data and analyses of other functional assessment scales/instruments. Rasch analysis of the data showed the items work well together to measure the concept of self-care, with generally good infit and outfit statistics. As expected, for each item, the average self-care ability Rasch measure of patients increases as the rating scale scores increase. All these results support the validity of the self-care data elements and scale in measuring self-care functional abilities.

Our analyses that focused on whether or not an IRF obtained The Joint Commission's Disease Specific Certification for Stroke Rehabilitation showed that IRFs with higher (better) computed performance measure scores were more likely to have this structural measure of quality (certification). These analyses support the validity of the calculated performance measure scores.

**2b2. EXCLUSIONS ANALYSIS**

**NA ☐ no exclusions — *skip to section*** 2b3

**2b2.1. Describe the method of testing exclusions and what it tests** (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

We examined the number and percentage of patients who were excluded from the performance measure calculation due to exclusion criteria. The exclusion criteria are applied to the data in order to maintain the validity of the calculated performance measure scores and were identified in consultation with expert panel members and in response to public comments. Some IRFs specialize in the care of patients with complex needs, for example, patients with traumatic spinal cord injury and traumatic brain injury; therefore, application of these exclusion criteria is important to ensure the validity of the calculated performance scores for all IRFs, regardless of whether the IRF offers specialized services for complex patients. All exclusion criteria were applied prior to our developing the risk-adjustment model.

For several exclusion criteria, the rationale for the exclusion of these patients is that improvement in self-care would be limited or unpredictable. For these exclusion criteria, we report the mean, median and 25$^{th}$ and 75$^{th}$ percentiles for change in self-care scores.

For patients who have an incomplete stay (e.g., emergency discharge), it is challenging to collect accurate discharge functional status data due to the urgent nature of the discharge. Therefore, patients with incomplete stays are excluded from the performance measure calculation, and we are unable to conduct analyses due to the unavailability of data. A total of 55,590 (11.3%) of patient stays were classified as incomplete stays based on the definition of an incomplete stay.

We excluded patients younger than 21 in our original measure specifications because we had very few patients in our sample younger than 21 and there is limited literature about functional outcomes for Medicare patients younger than 21. We are maintaining this exclusion criterion because there is still limited evidence in the literature about function outcomes for Medicare beneficiaries who are younger than 21 and there were only 32 patients younger than 21 discharged in calendar year 2017.

**2b2.2. What were the statistical results from testing exclusions**? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

A total of 65,017 patient stays (13.2%) are excluded from the change in self-care performance measure. As indicated above, most of these (55,590 (11.3%)) are due to incomplete stays. An analysis of differences between patient-level characteristics for those included and excluded from the performance measure (available upon request) show very little variation in the two populations. The largest difference was 1.1% and observed for gender (53.0% and 54.1% identified as female for the full population and the population with exclusions applied, respectively). As noted above, these exclusion criteria are important to apply to ensure the validity of the calculated performance scores for all IRFs, regardless of whether the IRF offers specialized services for complex patients.

**Table 11** shows the number and percent of patients excluded for each exclusion criteria, and the mean, median and 25$^{th}$ and 75$^{th}$ percentile for the change in self-care scores (values are reported as units of change in self-care, possible range: -35 to 35). For patients with persistent vegetative state, locked-in syndrome, those discharged to hospice and patients who are independent with all self-care activities on admission, analyses show these patients had very limited improvement with the self-care activities. For patients in a coma and those with severe brain damage, severe anoxic brain damage, and cerebral edema, improvement in self-care showed variability when we examined unadjusted data by quarter.

**Table 11. Observed Change in Self-Care Score in Self-Care Units by Exclusion Criteria* (N=493,209)**

| Exclusion Criteria | n (%) | Mean | SD | Median | 25th Percentile | 75th Percentile |
|---|---|---|---|---|---|---|
| All Excluded Medical Conditions | 7,650 (1.6) | 8.8 | 8.1 | 9.0 | 4.0 | 14.0 |
| Coma | 65 (<0.1) | 8.6 | 7.0 | 9.0 | 3.0 | 14.0 |
| Complete Tetraplegia | 311 (0.1) | 4.7 | 6.8 | 3.0 | 0.0 | 8.0 |
| Persistent vegetative state** | < 11 (<0.1) | ** | ** | ** | ** | ** |
| Severe brain damage | 731 (0.1) | 9.1 | 7.7 | 9.0 | 4.0 | 14.0 |
| Locked-In Syndrome | 12 (<0.1) | 3.6 | 6.7 | 0.0 | 0.0 | 6.0 |
| Severe anoxic brain damage, cerebral edema, or compression of the brain | 6,631 (1.3) | 8.9 | 8.1 | 9.0 | 4.0 | 14.0 |
| Discharged to Hospice | 2,548 (0.5) | 2.2 | 7.8 | 1.0 | -2.0 | 7.0 |
| Independent with all Admission Self-Care Data Elements | 714 (0.1) | -2.8 | 7.1 | 0.0 | -1.0 | 0.0 |

Note: N = number of patient stays; Observed Change in Self-Care values are reported as units of change in self-care (possible range: -35 to 35)
*For patients who have an incomplete stay (e.g., emergency discharge), it is challenging to collect accurate discharge functional status data. Therefore, we are unable to conduct analyses due to the unavailability of IRF-PAI data. For the exclusion criterion age younger than 21, we have not conducted analyses due to the very small number of patients in this age group. In calendar year 2017, there were 32 patients younger than 21.
**The number of patients with this medical condition is less than 11, and thus too small to publicly report.
Source: RTI analysis of IRF-PAI, January – December 2017. (Program reference: MV47)

**2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results?** (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

In calendar year 2017 data, 13.2% of patient stays were excluded from the calculated performance scores. The exclusion criteria are applied to the data in order to maintain the validity of the calculated performance measure scores. Data analysis results support these exclusions, because inclusion of limited and less predictable self-care improvement for these patients could affect computed performance measure scores for the selected IRFs that admit patients who meet these criteria.

---

**2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES**

***If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section*** 2b4**.**

**2b3.1. What method of controlling for differences in case mix is used?**

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with 93 risk factors**

☐ **Stratification by risk categories**

☐ **Other,**

**2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.**

The risk adjustment model, including the intercept (constant), covariates (risk factors) with definitions and coefficients are provided as an attached excel file and in Appendix A Table A-1. We used a Generalized Linear Model regression analysis to obtain the regression intercept (constant) and regression coefficients values.

***Model for Individual Patient's Expected Change in Self-Care Score***

The risk-adjustment model includes a total of 93 covariates. For each individual patient, not every covariate will apply, because, for example, only one age group, one primary diagnosis group, and one bladder incontinence covariate will apply. In addition, patients could have 0 or up to 40 comorbidities. Therefore, for an individual patient stay, up to 61 covariates may apply.

As described in the measure calculation algorithm, the regression intercept and coefficients are used to calculate an expected change in self-care score for each patient stay using the formula below:

*Expected change in self-care score =*

> intercept + (age group * coefficient) + (continuous admission self-care * coefficient) + (squared admission self-care * coefficient) + (primary diagnosis group * coefficient) + (interaction term for admission self-care and primary diagnosis group * coefficient) + (prior surgery * coefficient) + (prior functioning: self-care * coefficient) + (prior functioning: indoor ambulation * coefficient) + (prior use of walker * coefficient) + (prior use of wheelchair * coefficient) + (prior use of mechanical lift * coefficient) + (prior use of orthotics/prosthetics * coefficient) + (stage 2 pressure ulcer * coefficient) + (stage 3, 4 or unstageable pressure ulcer * coefficient) + (cognitive function * coefficient) + (communication impairment * coefficient) + (bladder incontinence * coefficient) + (bowel incontinence * coefficient) + (swallowing ability: modified food consistency * coefficient) + (swallowing ability: tube/parenteral feeding * coefficient) + (low BMI * coefficient) + (comorbidity * coefficient)

In the equation above, the intercept and coefficient values were constant for each patient, while risk adjustor values were specific to the patient. Patients could have multiple comorbidities.

We provide detailed measure calculation instructions for this performance measure in an attachment in the "NQF Specifications" document. The detailed measure calculation instructions are available to the public in the document entitled "IRF Quality Reporting Program Measure Calculations and Reporting User's Manual" that can be found at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/IRF-Quality-Reporting/IRF-Quality-Reporting-Program-Measures-Information-.html. The current version of the manual, Version 3.0, reflects current measure specifications. We will implement the changes described in the NQF specifications document in 2019.

***Risk Adjusted Change in Self-Care Outcome for Each IRF***

To calculate the risk adjusted change score for each IRF, we compute three values:

1. **Mean observed change in self-care score for each IRF:** We calculated the mean observed change score for each IRF as the mean of the observed change in self-care scores for all patients in the IRF.
2. **Mean expected change in self-care score for each IRF:** We calculate each patient's expected change in self-care score using the intercept and coefficient values that apply to the patient. We then compute the mean expected change in self-care score for each IRF by calculating the mean of the expected change score for all patients in the IRF.
3. **National mean observed change in self-care score:** We calculated national mean observed change in self-care score using data for all patients and all IRFs.

Using the above three values, the risk adjusted change in self-care outcome for each IRF is calculated using the formula:

$$(\text{IRF mean observed change score} - \text{IRF mean expected change score}) + \text{National mean observed change score}$$

**2b3.2. If an outcome or resource use component measure is <u>not risk adjusted or stratified</u>, provide <u>rationale and analyses</u> to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities**.

Not applicable. This performance measure is risk adjusted.

**2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk** (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

This performance measure estimates the risk adjusted mean change in self-care score between admission and discharge among IRF patients. Functional improvement can vary based on patients' demographic or clinical characteristics, therefore, this measure is risk adjusted. The goal of risk adjustment is to control for differences across facilities in patient characteristics at admission that might be related to the outcome of interest. This allows outcomes to be compared across facilities after differences in patient complexity (i.e., patient characteristics) have been accounted for in the analysis. The risk adjustment model for this measure controls for variation across facilities in patient demographics (e.g., age) and clinical characteristics (e.g., diagnosis) present at the time of admission that may influence functional outcomes, to allow change in self-care outcomes to be compared across IRFs.

Initial development of the risk adjustment model can be found on this measure's previous testing form. We are now updating the risk adjustment model for this measure using the national data collected using the Inpatient Rehabilitation Facility Patient Assessment Instrument (IRF-PAI) including patients' primary conditions, prior functioning, and comorbidities at admission. Testing of the risk adjustment model was conducted after applying the exclusion criteria described in 2b2.

***Risk Adjustor Selection – Conceptual Rationale and Statistical Testing***

The initial selection of risk adjustors was based on a review of the literature, input from technical experts and public comments, followed by data analysis. Please see the 2014 testing form on this measure for more detailed information on the initial selection of risk adjustors for this measure. In preparation for endorsement maintenance, we updated our literature review and conducted additional analyses.

We tested the risk adjustors using a generalized linear model with generalized estimation equations (GEE) as the estimation method to account for clustering of data within each IRF. The generalized estimation equations method accounted for potentially correlated outcomes of patients within the same IRF, in addition to risk adjusting the change in self-care outcome using the final set of risk adjustors.

The dependent variable was the change in self-care score for each patient, calculated as the difference between the discharge self-care score and admission self-care score. The regression coefficient represents the effect of an individual covariate. For example, a coefficient value of -0.5 for a comorbidity would be interpreted to mean that, on average, patients with that comorbidity had a change in self-care score that was 0.5 self-care units less than patients without that comorbidity.

Risk adjustors were added to the model together and decisions were made to retain or drop each risk adjustor based on its sample size, regression coefficient, significance level, and clinical relevance to self-care outcomes. For example, we dropped comorbidities that no longer showed a negative association with the dependent variable, because comorbidities are expected to limit functional improvement. We added comorbidities that showed a significant negative association with the dependent variable. The final risk adjustor decisions were based on a combination of clinical reasoning and statistical findings.

Risk adjustors included in the final model are described below, and also presented in S. 2b. Data Dictionary, Code Table, or Value Sets.

***Age Groups:*** We included seven age groups in the risk adjustment model (< 35 years, 35–44 years, 45–54 years, 55–64 years, 75–84 years, 85–90 years, and ≥ 90 years). The age group 65–74 years formed the reference category. Age was not normally distributed in our sample, so it was more appropriate to use age groups in our analyses. When compared to the reference group (patients 65–74 years), patients 35–44 years (coefficient = 0.2755, $p = 0.002$), 45–54 years (coefficient = 0.4569, $p < 0.001$), and 55–64 years (coefficient = 0.2666, $p < 0.001$) had significantly larger change in self-care scores. Patients 75–84 years (coefficient = -0.5869, $p < 0.001$), 85–90 years (coefficient = -1.2785, $p < 0.001$), and over 90 years (coefficient = -2.0677, $p < 0.001$) had significantly, and progressively, smaller change in self-care scores than patients in the reference category. Patients younger than 35 years did not have significantly different change scores compared with the reference category. Nevertheless, we chose not to collapse this group based on public comment feedback regarding the clinical importance of maintaining fine discrimination among age groups.

***Admission Self-Care Scores:*** Since improvement in self-care during the IRF stay may vary based on admission self-care ability, we risk adjusted for admission self-care scores in our regression model. Scatter plots of admission self-care scores and change in self-care scores showed a non-linear relationship between the two variables. Therefore, we included admission self-care scores in two forms in the model: a continuous form, and a squared form to account for the curvilinear relationship. The continuous form (coefficient = 0.0793, $p < 0.001$) and the squared form (coefficient = -0.0163, $p < 0.001$) of admission self-care scores had significant effects. Thus, we included both forms of admission self-care in the final model.

***Primary Diagnosis Groups Based on IRF Primary Diagnosis:*** We used Impairment Group codes reported on the IRF-PAI (Item 21) to create the following 13 mutually-exclusive primary diagnosis groups: (1) stroke, (2) non-traumatic brain dysfunction, (3) traumatic brain dysfunction, (4) non-traumatic spinal cord dysfunction, (5) traumatic spinal cord dysfunction, (6) progressive neurological conditions, (7) other neurological conditions (e.g., polyneuropathy), (8) fractures and other multiple trauma, (9) hip and knee replacements, (10) amputation, (11) other orthopedic conditions (e.g., arthritis), (12) debility and cardiorespiratory conditions, and (13) medically complex conditions. “Hip and knee replacements” formed the reference category, and the remaining 12 primary diagnosis groups were risk adjustors in the model. When compared to the reference category, all diagnosis groups were significant predictors of change in self-care scores. The primary diagnosis groups had significantly smaller change in self-care scores compared with the “hip and knee replacements” group. The “traumatic spinal cord dysfunction” group had the largest coefficient (-10.1155, $p < 0.001$).

***Interaction between Primary Diagnosis Groups and Admission Self-Care Scores:*** The relationship between admission self-care and change in self-care scores may vary based on the patient’s primary diagnosis group. To account for this, we tested interaction terms between admission self-care scores (continuous form) and each diagnosis group included in the model. Thus, 12 interaction terms for admission self-care by diagnosis group were tested. All interaction terms were significant, as shown in Attachment 1. All interaction terms were retained in the model.

***Prior Surgery:*** We included patients who had a major surgery during the 100 days prior to admission as a risk adjustor in the model, because patients who have recently undergone a major surgery tend to have more functional improvement than patients with medical issues without surgery (Coefficient = 0.1952, $p < 0.001$).

***Prior Functioning - Self-Care:*** We included patients’ functional ability in self-care before the onset of their current illness, injury or exacerbation, as a risk adjustor in the model. We included separate categories for patients who were “dependent”, and those who needed “some help” in self-care before their current medical issue. Patients who were “previously independent” in self-care formed the reference category. Patients who were previously “dependent” in self-care and those who needed “some help” had significantly smaller change scores compared with the reference category. The negative coefficient for the “dependent” category (coefficient = -4.3325, $p < 0.001$) was larger than that for the “some help” (coefficient = -1.9657, $p < 0.001$) category.

***Prior Functioning - Indoor Ambulation:*** We included patients' functional ability in indoor ambulation before onset of their current illness, injury or exacerbation, as a risk adjustor in the model. We combined the "dependent" and "some help" categories into one group, and patients who were previously independent in indoor ambulation formed the reference category. Patients who were previously dependent or needed some help in indoor ambulation had significantly smaller change in self-care scores (coefficient = -0.8074, $p < 0.001$) compared with the reference category.

***Prior Mobility Devices/Aids:*** We risk adjusted for use of four types of mobility devices or aids before the current illness, injury, or exacerbation, including walker, wheelchair/scooter (full time/part time), mechanical lift, and orthotics/prosthetics. Prior use of each of these mobility devices or aids was associated with significantly smaller change in self-care scores, with prior use of a mechanical lift having the largest coefficient (coefficient = -1.9134, $p < 0.001$), followed by prior use of wheelchair or scooter (coefficient = -0.9244, $p < 0.001$), prior use of orthotics/prosthetics (-.4816, $p < 0.001$) and prior use of a walker (coefficient = -0.0560, $p = 0.004$).

***Stage 2 Pressure Ulcer****:* Our risk adjustment model included an indicator variable for the presence of one or more stage 2 pressure ulcers on admission, with the reference category being patients who did not have a stage 2 pressure ulcer. Patients with stage 2 pressure ulcers had a significantly smaller change in self-care scores (coefficient = -0.9422, $p < 0.001$) compared with the reference category.

***Stage 3, 4, or Unstageable Pressure Ulcers*** We included an indicator variable for the presence of one or more stage 3, 4, or unstageable pressure ulcers, with the reference category being patients who did not have such ulcers. Patients with stage 3, 4, or unstageable pressure ulcers had significantly smaller change in self-care scores (coefficient = -1.3024, $p < 0.001$) compared with the reference category.

***Cognitive Function Assessed by the Brief Interview for Mental Status:*** Based on Brief Interview for Mental Status (BIMS) scores, patients' cognitive function was classified as intact or borderline, moderately impaired, or severely impaired. "Moderately impaired" and "severely impaired" cognitive function were included as two separate risk adjustors in the model, while "intact or borderline" cognitive function formed the reference category. Patients with moderately impaired cognitive function (coefficient = -0.7763, $p < 0.001$) and those with severely impaired cognitive function (coefficient = -1.7718, $p < 0.001$) had significantly smaller change scores compared with the reference category.

***Communication Impairment:*** Communication impairment includes both expression (expression of ideas and wants) and comprehension (understanding verbal content) abilities. While expression and comprehension abilities are separate assessment items, we combined them into a single communication impairment risk adjustor given these two variables were correlated, with considerable overlap in patients who had expression and comprehension impairment and based on input from the expert panel. The final risk adjustment model included "moderate to severe communication impairment" as a risk adjustor, with this group having significantly smaller change in self-care scores (coefficient = -0.9358, $p < 0.001$) compared with the reference category of "mild or no communication impairment."

***Bladder Incontinence:*** We included a single risk adjustor for bladder incontinence, which comprises patients with bladder incontinence "less than daily," "daily," or "always." The reference category included patients who had "stress incontinence only, were always continent, or had no urine output (e.g., renal failure)." Patients with bladder incontinence (coefficient = -1.2459, $p < 0.001$) had significantly smaller change in self-care scores compared with the reference category. We also included a risk adjustor for patients with an indwelling urinary catheter (coefficient = -1.1527, $p < 0.001$).

***Bowel Incontinence:*** We included two separate risk adjustors related to bowel incontinence: "always incontinent" and "less than daily or daily incontinence." The reference category included patients who "were always continent, had no bowel output during the assessment period, or had a bowel catheter management system". Patients with bowel incontinence had significantly smaller change in self-care scores compared with the reference group, with the "always incontinent" category (coefficient = -2.3437, $p < 0.001$) having a larger

negative coefficient compared with the "less than daily" or "daily incontinence" category (coefficient = -0.8470, $p < 0.001$).

***Swallowing Ability:*** Our model included two separate risk adjustors related to patients' swallowing ability: (1) need for modified food consistency, and (2) need for tube or parenteral feeding. Both risk adjustors were significantly predictive of smaller change in self-care scores, with the tube or parenteral feeding group (coefficient = -0.7813, $p < 0.001$) having a larger coefficient than the modified food consistency group (coefficient = -0.6785, $p < 0.001$).

***Low Body Mass Index (BMI):*** We included a risk adjustor for patients with low BMI based on their height and weight. Patients with low BMI had significantly smaller change in self-care scores (coefficient= -0.3363, $p < 0.001$) compared with the reference category.

***Comorbidities:*** We used all the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) codes reported on the IRF-PAI (Item 24 - Comorbid Conditions) to identify patient comorbidities. ICD-10-CM codes were used to assign patients into one or more of the Hierarchical Condition Categories. We tested approximately 135 of the Hierarchical Condition Categories that were determined to be clinically relevant to self-care outcomes.

To ensure that the same diagnoses or conditions were not represented in both the primary diagnosis groups and comorbidities, we applied exclusion criteria such that certain comorbidities were excluded if they were also present as primary diagnoses. For example, tetraplegia and paraplegia were excluded as comorbidities if the patient's primary diagnosis group was "non-traumatic spinal cord dysfunction" or "traumatic spinal cord dysfunction"; amputation was excluded as a comorbidity if the patient's primary diagnosis group was "amputation."

The attached model in S.2b. Data Dictionary, Code Table, or Value Sets shows the regression coefficients and significance values for all comorbidities in the final risk adjustment model. We retained comorbidities that were clinically important or had large coefficients, even when they were not statistically significant. Comorbidities with the largest negative coefficients, indicating smaller change in self-care scores, include certain cancers; dementia; tetraplegia; Amyotrophic Lateral Sclerosis and other motor neuron diseases; Cerebral Palsy; legally blind; and major fracture, except of skull, vertebrae, or hip.

**2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:**

☐ **Published literature**

☒ **Internal data analysis**

☐ **Other (please describe)**

**2b3.4a. What were the statistical results of the analyses used to select risk factors?**

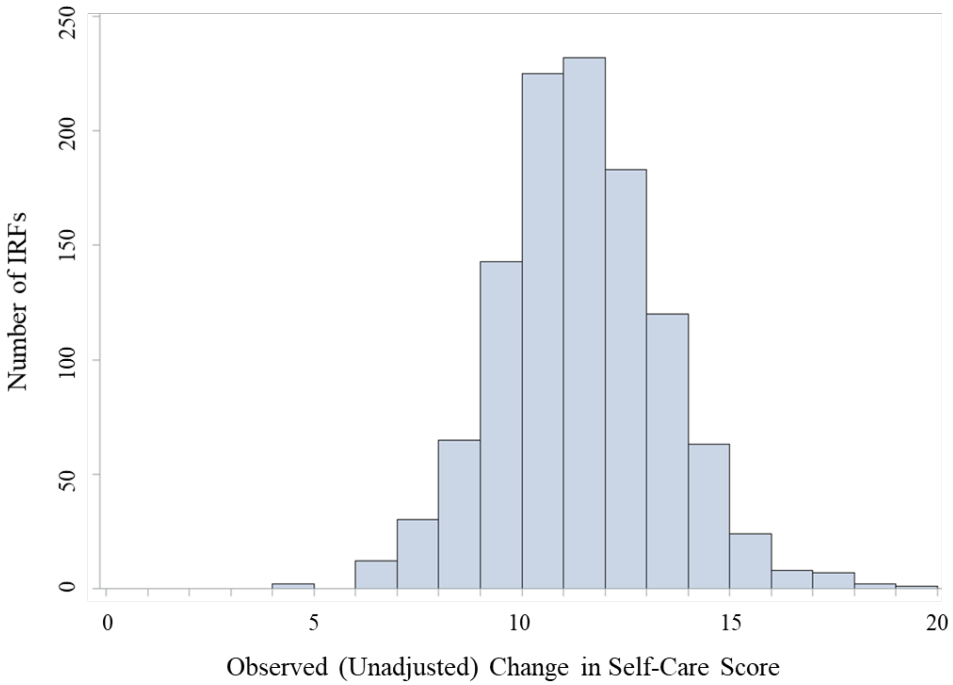
Results of the final risk adjustment model are shown in S.2b. Data Dictionary, Code Table, or Value Sets, along with regression coefficients and significance values of the final set of risk adjustors.

As described above, decisions were made to retain or drop each risk adjustor based on its sample size, regression coefficient, significance level, and clinical relevance to self-care outcomes. For example, we dropped comorbidities that no longer showed a negative association with the dependent variable, because comorbidities are expected to limit functional improvement. We added comorbidities that showed a significant negative association with the dependent variable. The final risk adjustor decisions were based on a combination of clinical reasoning and statistical findings.

The overall model was a significant predictor of change in self-care scores, with a *p*-value less than 0.001. The overall model R-square was 0.23, indicating that 23% of the variance in change in self-care was explained by the model. In general, regression coefficients of individual risk adjustors demonstrated that the predictive ability of risk adjustors was as clinically expected.

Distributions of the facility-level mean unadjusted and risk adjusted change in self-care scores are shown in **Figures 2 and 3 and Table 12.** Important differences in the distribution of the two sets of scores were noted, which speaks to the importance of risk adjustment. **Figure 2** shows that the facility-level mean unadjusted change scores are largely concentrated in the center of the distribution, with fewer IRFs at the extremes of the distribution, particularly at the higher extreme. In contrast, **Figure 3** demonstrates normal distribution and good variability of the facility-level mean risk adjusted change in self-care scores.

**Figure 2. Distribution of Facility-Level Mean Unadjusted Change in Self-Care Scores (n=1,117)**




Source: RTI analysis of IRF-PAI, January – December 2017. (Program reference: MV50)

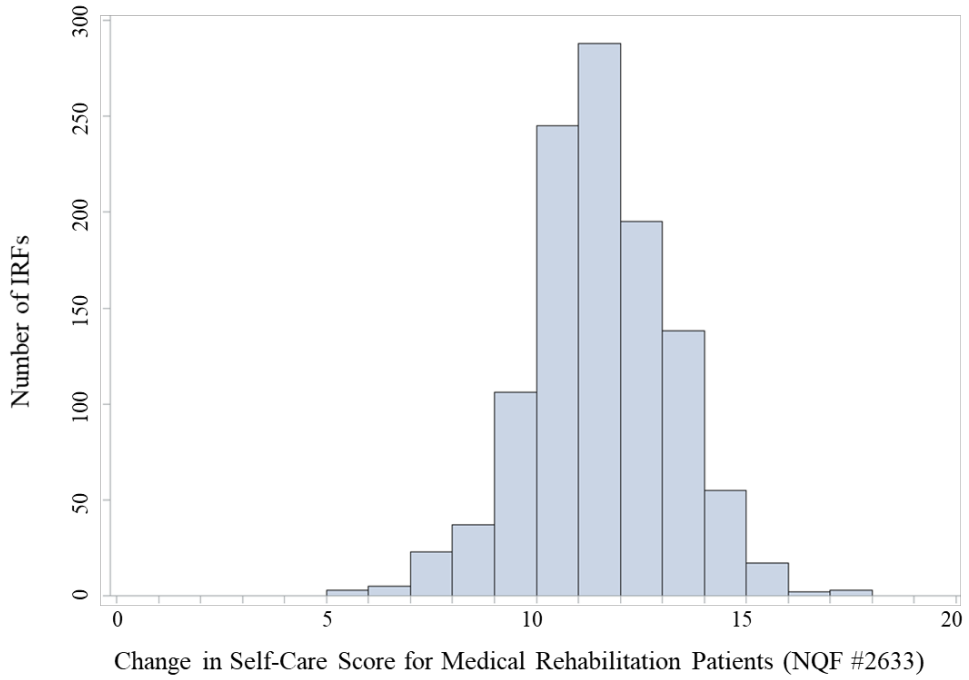
**Figure 2 Data Table. Distribution of Facility-Level Mean Unadjusted Change in Self-Care Scores (n=1,117)**

| Observed (Unadjusted) Change in Self-Care Scores* | Number of IRFs |
|---|---|
| 4.0 | 2 |
| 6.0 | 6 |
| 7.0 | 20 |
| 8.0 | 45 |
| 9.0 | 93 |
| 10.0 | 178 |
| 11.0 | 251 |
| 12.0 | 231 |
| 13.0 | 132 |
| 14.0 | 91 |
| 15.0 | 45 |
| 16.0 | 9 |
| 17.0 | 9 |
| 18.0 | 4 |
| 20.0 | 1 |

| Observed (Unadjusted) Change in Self-Care Scores* | Number of IRFs |
|---|---|
| Total | 1117 |

*Scores were rounded to the nearest whole number for the figure
Source: RTI analysis of IRF-PAI, January – December 2017. (Program reference: MV50)

**Figure 3. Distribution of Facility-Level Mean Risk Adjusted Change in Self-Care Scores (n=1,117)**




Source: RTI analysis of IRF-PAI, January – December 2017. (Program reference: MV50)

**Figure 3 Data Table. Distribution of Facility-Level Mean Risk Adjusted Change in Self-Care Scores (n=1,117)**

| Facility-Level Mean Risk Adjusted Change in Self-Care Scores* | Number of IRFs |
|---|---|
| 5.0 | 2 |
| 6.0 | 3 |
| 7.0 | 11 |
| 8.0 | 34 |
| 9.0 | 63 |
| 10.0 | 179 |
| 11.0 | 266 |
| 12.0 | 262 |
| 13.0 | 159 |
| 14.0 | 98 |
| 15.0 | 27 |
| 16.0 | 8 |
| 17.0 | 4 |
| 18.0 | 1 |
| **Total** | **1,117** |

*Scores were rounded to the nearest whole number for the figure
Source: RTI analysis of IRF-PAI, January – December 2017. (Program reference: MV50)

Skewness and kurtosis values of the facility-level mean unadjusted change scores are larger than those of the mean risk adjusted change scores (**Table 12**), indicating that the unadjusted scores deviate from a normal distribution to a larger extent than the risk adjusted scores (i.e., performance measure scores). **Table 12** shows that the standard deviation of the mean risk adjusted change scores is slightly smaller than those of the unadjusted change scores. The mean risk adjusted change scores have a range of 5.4 to 17.5, and an interquartile range of 2.20. In contrast, the mean unadjusted change scores have a wider range of 4.2 to 19.6, and an interquartile range of 2.41.

**Table 12. Distribution of Facility-Level Mean Unadjusted and Risk Adjusted Change in Self-Care Scores (n=1,117)**

| Change in Self-Care Score | N | Mean (SD) | SE | Min | 10th Pctl | 25th Pctl | Median | 75th Pctl | 90th Pctl | Max | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Unadjusted (Observed) | 1,117 | 11.4 (2.0) | 0.1 | 4.2 | 9.1 | 10.1 | 11.3 | 12.6 | 14.0 | 19.6 | 0.2 | 0.7 |
| Risk Adjusted Performance Measure | 1,117 | 11.5 (1.7) | 0.1 | 5.4 | 9.4 | 10.4 | 11.5 | 12.6 | 13.7 | 17.5 | -0.1 | 0.6 |

N = Number; SD = Standard deviation; SE = standard error; Min = Minimum; Pctl = Percentile; Max = Maximum
Note: Providers with <20 stays during the 12-month testing period are excluded from facility-level analyses.
Source: RTI analysis of IRF-PAI, January – December 2017. (Program reference: MV50)

**2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors** *(e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

We examined whether 5 social risk factors affected computed performance measure scores: 1) dual eligibility (patient-level variable); 2) race/ethnicity (patient-level variable); 3) living alone (patient-level variable); 4) urbanicity based on the patient's residence (community-level variable); and 5) socioeconomic status (SES) (community-level variable).

We obtained patients' dual eligibility status from the Integrated Data Repository (IDR), and race/ethnicity and living alone status from the IRF-PAI. Urbanicity was determined by cross-walking beneficiary residence ZIP codes (from the IRF-PAI) to Federal Information Processing Standard Publication (FIPS) codes,[4] then cross-walking FIPS codes to Rural-Urban Commuting Area Codes (RUCA_2013).[5] Socioeconomic status was determined using the Agency of Healthcare Research and Quality's SES Index[6] calculated based on beneficiary residence ZIP Code Tabulation Area (ZCTA). ZCTA was found by cross-walking the beneficiary residence ZIP code with ZCTA. We used data from the 2016 American Community Survey (5-year file) to calculate AHRQ SES Index, with higher values indicating higher SES.

We conducted the following analyses to examine the effect of the 5 social risk factors:

- We calculated the percentage of stays for each social risk factor subgroup;
- We calculated the change in self-care score for each social risk factor subgroup;
- We added indicators for each social risk factor to our risk adjustment model and estimated the coefficients of these risk adjusters in the model; and

[4] https://www.huduser.gov/portal/datasets/usps_crosswalk.html

[5] https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes.aspx

- We calculated the difference in provider scores with and without social risk factor adjustment.

**Table 13** shows the distribution of the social risk factors in the calendar 2017 IRF data and the mean change in self-care score by social risk factor subgroup. We found that 12.2% of patients were dual eligible with full Medicaid benefits, 79.4% of patients were white, and 29.7% were living alone. We also found that 83.8% of IRF patients lived in urban areas. The lowest quartile of AHRQ SES index ranged from 27.9 - 49.5; the highest quartile ranged from 55.3 – 75.7.

The mean unadjusted change in self-care score varied by dual eligibility status, race, and living alone status. Patients who were dual eligible with full Medicaid benefits had on average 11.0 units of change in self-care while patients who were dual eligible without full Medicaid benefits or who were non-dual eligible had on average 12.0 and 11.6 units of change in self-care, respectively. For race, the highest mean change in self-care during 2017 was found among patients who were white (11.6 units of change), multiracial (11.5 units of change), or Native American or Alaskan Native (11.4 units of change) whereas the lowest was among patients who were Asian (10.4 units of change). Patients who were living alone prior to their hospitalization had on average 12.0 units of change in self-care whereas those not living alone had 11.3 units of change in self-care. The mean unadjusted change in self-care scores were similar across Hispanic ethnicity, urbanicity, and SES. Patients who were Hispanic had a similar mean change in self-care score (11.8 units of change) as patients who were of non-Hispanic ethnicity (11.5 units of change). The average unit of change in self-care for patients who were living in rural and urban locations, ranged from 11.5 to 11.7 units of change in self-care, and by AHRQ SES Index, ranged from to 11.2 units of change for quartile 4 to 11.7 units of change for quartile 1.

**Table 13. Distribution of Social Risk Factors and Mean Change in Self-Care Score for IRF Patients (N = 428,710)**

| Social Risk Factor | n (%) | Observed Change in Self-Care (unadjusted) |
|---|---|---|
| ***Dual Eligibility*** | | |
| Dual with full Medicaid | 52,450 (12.2) | 11.0 |
| Dual without full Medicaid | 25,113 (5.9) | 12.0 |
| Non-dual | 351,147 (81.9) | 11.6 |
| ***Race*** | | |
| White | 340,398 (79.4) | 11.6 |
| Black | 46,949 (11.0) | 10.9 |
| Asian | 6,689 (1.6) | 10.4 |
| American Indian or Alaska Native | 1,339 (0.3) | 11.4 |
| Native Hawaiian or Pacific Islander | 1,546 (0.4) | 10.7 |
| Multiracial | 246 (0.1) | 11.5 |
| Missing | 31,543 (7.4) | 11.8 |
| ***Hispanic Ethnicity*** | | |
| Yes | 20,147 (4.7) | 11.8 |
| No | 408,563 (95.3) | 11.5 |
| ***Living Alone*** | | |
| Yes | 127,218 (29.7) | 12.0 |
| No | 301,492 (70.3) | 11.3 |
| ***Urbanicity*** | | |
| Urban | 359,388 (83.8) | 11.5 |
| Suburban | 48,965 (11.4) | 11.8 |
| Rural | 18,000 (4.2) | 11.7 |
| Missing | 2,357 (0.5) | 10.9 |

| Social Risk Factor | n (%) | Observed Change in Self-Care (unadjusted) |
|---|---|---|
| ***AHRQ SES Index**** | | |
| Quartile 1 (27.9 - 49.5) | 106,256 (24.8) | 11.7 |
| Quartile 2 (49.5 – 52.1) | 106,438 (24.8) | 11.7 |
| Quartile 3 (52.1 – 55.3) | 106,876 (24.9) | 11.5 |
| Quartile 4 (55.3 – 75.7) | 107,203 (25.0) | 11.2 |
| Missing | 1,937 (0.5) | 10.8 |

* based on beneficiary residence. AHRQ = Agency for Healthcare Research.

Notes: N= number of patient stays; patient-level exclusion criteria applied; unadjusted Change in Self-Care values are reported as units of change in self-care (possible range: -35 to 35).
Source: RTI analysis of IRF-PAI, January – December 2017. (Program reference: LP65)

**Table 14** shows the social risk factor estimates in our Generalized Linear regression model. Dual eligibility patients with full Medicaid benefits had lower self-care changes scores (coefficient = -0.3760, $p < 0.001$) while patients with partial Medicaid benefits had higher self-care change scores (coefficient =0.2438, $p < 0.001$), on average, than patients who were non-dual eligible. Compared to patients who were White, Black patients (coefficient = -0.5551, $p < 0.001$), Asian patients (coefficient = -0.4772, $p < 0.001$), and patients of Native Hawaiian or Pacific Islander descent (coefficient = -0.4248, $p = 0.003$) had lower self-care changes, on average. Hispanic patients (coefficient = 0.1771, $p = 0.004$) had higher self-care change scores than non-Hispanic patients. Patients who lived alone (coefficient = 0.4732, $p < 0.001$) had higher self-care change scores than patients who did not live alone prior to their hospitalization. Patients living in rural areas had similar change in self-care scores compared with patients living in urban areas. Patients residing in AHRQ SES Index quartiles 1-3 had higher self-care change scores, on average, than patients residing in AHRQ SES Index quartile 4.

**Table 14. Effect of Social Risk Factors in the IRF Change in Self-Care Regression Model (N = 428,192)**

| Social Risk Factor | Estimate | SE | *p*-value |
|---|---|---|---|
| ***Dual Eligibility (reference = Non-dual)*** | | | |
| Dual with full Medicaid | -0.3760 | 0.03 | <.001 |
| Dual without full Medicaid | 0.2438 | 0.04 | <.001 |
| ***Race/Ethnicity (reference = White)*** | | | |
| Black | -0.5551 | 0.03 | <.001 |
| Asian | -0.4772 | 0.07 | <.001 |
| American Indian or Alaska Native | -0.1013 | 0.16 | 0.514 |
| Native Hawaiian or Pacific Islander | -0.4248 | 0.14 | 0.003 |
| Multiracial | 0.1038 | 0.36 | 0.774 |
| Missing | -0.0252 | 0.05 | 0.610 |
| ***Hispanic Ethnicity*** | 0.1771 | 0.06 | 0.004 |
| ***Living Alone*** | 0.4732 | 0.02 | <.001 |
| ***Urbanicity* (reference = Urban)*** | | | |
| Rural | -0.0466 | 0.04 | 0.292 |
| Suburban | 0.0296 | 0.03 | 0.298 |
| Missing | -0.2512 | 0.28 | 0.361 |
| ***AHRQ SES Index* (reference = Quartile 4 (55.6 to 75.7))*** | | | |
| Quartile 1 (28.9 to 49.6) | 0.5516 | 0.03 | <.001 |
| Quartile 2 (49.7 to 52.2) | 0.3961 | 0.03 | <.001 |
| Quartile 3 (52.3 to 55.5) | 0.2917 | 0.02 | <.001 |

| Social Risk Factor | Estimate | SE | *p*-value |
|---|---|---|---|
| Missing | 0.0140 | 0.75 | 0.985 |

* based on patient residence. AHRQ = Agency for Healthcare Research.

Note: SE=Standard error; Patient-level exclusion criteria applied.

Source: RTI analysis of IRF-PAI, January – December 2017. (Program reference: LP65)

**Table 15** shows the distribution of the change in self-care performance measure scores with and without social risk factor adjustment. Overall, social risk factor adjustment had minimal impact on providers' calculated performance measure scores. Between the two sets of scores, the difference in mean scores was 0.0 units of change in self-care, with a standard deviation of 0.2 and interquartile range of 0.2.

**Table 15: Distribution of IRF Change in Self-Care Scores With and Without Adjustment for Social Risk Factors (n = 1,117)**

| Change in Self-Care Scores | Mean | SD | Min | 25th Pct | Median | 75th Pct | Max |
|---|---|---|---|---|---|---|---|
| Not adjusting for SRF | 11.5 | 1.7 | 5.4 | 10.4 | 11.5 | 12.6 | 17.5 |
| Adjusting for SRF | 11.5 | 1.7 | 5.3 | 10.4 | 11.5 | 12.5 | 17.7 |
| Difference in Scores (SRF-adjusted minus non-SRF adjusted scores)* | 0.0 | 0.2 | -0.4 | -0.1 | 0.0 | 0.1 | 0.6 |

*Calculated as SRF-adjusted score minus non-SRF adjusted score for each facility.

Note: SD=Standard deviation; Min=minimum score; Max=maximum score; Pct = percentile. SRF = social risk factors.

Source: RTI analysis of IRF-PAI, January – December 2017. (Program reference: LP65)

Our testing of social risk factors and their relationship to patients' change in self-care scores indicate that some factors (full dual eligibility, Black, Asian or Native Hawaiian race) were tied to lower self-care change scores while others (lower SES, living alone, Hispanic ethnicity) were tied to higher self-care change scores. Although race and dual eligibility were associated with lower changes in self-care scores the effects were small, and we believe that further study is needed to better understand how social risk factors can influence health outcomes. Our risk adjustment model explained 23% of variance in change in self-care, and the inclusion of these five social risk factors did not explain any additional variance (r-squared = 0.234). In addition, the mean and median Change in Self-Care Score with and without adjusting for the social risk factors are the same.

As noted in the Assistant Secretary for Planning and Evaluation's Report to Congress entitled "*Social Risk Factors Performance under Value-Based Purchasing*" (https://aspe.hhs.gov/pdf-report/report-congress-social-risk-factors-and-performance-under-medicares-value-based-purchasing-programs), adjusting performance measures for social factors may mask disparities in the quality of care provided, which could reduce the ability to identify and reduce them. In addition, when differences in quality are related to poor performance, bias, or discrimination, adjusting performance measures could excuse the delivery of worse care to beneficiaries with social risk factors.

Therefore, we do not adjust for social risk factors in our risk adjustment model for the IRF Change in Self-Care performance measure. We will continue to monitor the impact of social risk factors on providers' performance measure scores.

**2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach** (*describe the steps—do not just name a method; what statistical analysis was used*)

*Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below*.

***If stratified, skip to*** 2b3.9

Our risk adjustment model demonstrates reasonable predictive validity for IRF change in self-care scores. Using multiple linear regression, we conducted regression diagnostics to assess model performance, examining predictive ability, and outlier influence.

**2b3.6. Statistical Risk Model Discrimination Statistics** (*e.g., c-statistic, R-squared*)**:**

Overall, the model explained 23% of variance in change in self-care.

**2b3.7. Statistical Risk Model Calibration Statistics** (*e.g., Hosmer-Lemeshow statistic*):

We conducted outlier influence analysis to assess for any outlying observations that may have large or extreme effects on the change in self-care outcome, with a Cook's D score of 1.0 or higher suggesting a potentially influential observation. All Cook's D scores were less than 1.0, with the maximum score being 0.0013.

**2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves**:

To assess model performance and stability across the sample, we divided our dataset into deciles of expected values and calculated the ratio of average expected change score to average observed change score within each decile. A ratio of 1 would indicate perfect agreement between average expected and observed change scores. We expect that the risk adjusted model performance will be stable among IRFs regardless of whether they have patients with low or high change scores on average.

As seen in **Table 16**, the average expected to observed change score ratios within each decile approximated 1.2, with a range of 0.9 to 1.3, validating model performance. There was little variability in average expected to observed change score ratios across deciles, supporting model stability across the range of expected change scores and across the sample.

**Table 16. Ratio of Average Expected to Average Observed Change in Self-Care Scores Across Deciles of Expected Change Scores**

| Deciles of Expected Change Scores | N | Average Expected Change Score | Average Observed Change Score | Average Expected to Observed Ratio |
|---|---|---|---|---|
| Decile 1 (-6.4 – 7.5) | 42,819 | 5.7 | 6.2 | 0.9 |
| Decile 2 (7.5 – 9.0) | 42,819 | 8.3 | 8.3 | 1.1 |
| Decile 3 (9.0 – 10.0) | 42,819 | 9.5 | 9.3 | 1.2 |
| Decile 4 (10.0 – 10.9) | 42,820 | 10.5 | 10.3 | 1.3 |
| Decile 5 (10.9 – 11.8) | 42,819 | 11.3 | 11.1 | 1.3 |
| Decile 6 (11.8 – 12.5) | 42,819 | 12.2 | 12.0 | 1.3 |
| Decile 7 (12.5 – 13.3) | 42,820 | 12.9 | 12.8 | 1.3 |
| Decile 8 (13.3 – 14.2) | 42,819 | 13.8 | 13.7 | 1.3 |
| Decile 9 (14.2 – 15.4) | 42,819 | 14.8 | 14.8 | 1.2 |
| Decile 10 (15.4 – 23.6) | 42,819 | 16.5 | 17.1 | 1.2 |
| **Total Sample** | 428,192 | 11.5 | 11.5 | 1.2 |

Note: N = number of patient stays; Observed Change in Self-Care values are reported as units of change in self-care (possible range: -35 to 35); Providers with < 20 stays during the 12-month testing period are excluded.
Source: RTI analysis of IRF-PAI, January – December 2017. (Program reference: MV50)

**2b3.9. Results of Risk Stratification Analysis**:

Not applicable – no stratification

**2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)?** (i.*e., what do the results mean and what are the norms for the test conducted*)

In summary, our results demonstrate reasonable predictive ability of our risk adjustment model for IRFs.

**2b3.11. Optional Additional Testing for Risk Adjustment** (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

None

_____

**2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE**

**2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified** (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

For the IRF Change in Self-Care Score performance measure, we examined whether each IRF's calculated performance measure score (i.e., the risk-adjusted change in self-care score) was worse than, better than, or the same as national average performance of all IRFs. For each IRF, we calculated the 95% confidence interval for the computed performance measure score and compared this with the national mean observed change score. Facilities whose confidence interval was lower than the national mean observed change score were considered to have worse performance than the national average. Facilities whose confidence interval was higher than the national mean observed change score were considered to have better performance than the national average. Facilities whose confidence interval overlapped with the national mean observed change score were considered to be similar to national average performance.

**2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities?** (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

**Table 17** shows that for the IRF Change in Self-Care Score measure, 33.0% of IRFs had 95% confidence intervals lower than the national mean change score, indicating worse than national average performance. As shown in **Figure 3** above, the IRF calculated performance scores (i.e., the risk-adjusted change in self-care scores) are generally normally distributed.

**Table 17. Comparison of Facility-Level Measure Scores with National Average Performance for Change in Self-Care Score (N = 1,117)**

| Measure Name | Facility Performance Worse than National Average N (%) | Facility Performance Better than National Average N (%) | Facility Performance Same as National Average N (%) |
|---|---|---|---|
| Change in Self-Care Score | 369 (33.0%) | 341 (30.5%) | 407 (36.4%) |

Note: Providers with < 20 stays during the 12-month testing period are excluded.

Source: RTI analysis of IRF-PAI, January – December 2017. (Program reference: MV53)

**2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities?** (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

These results demonstrate the ability of the measures to discriminate among facilities based on facility-level measure performance.

_____

**2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS**

***If only one set of specifications, this section can be skipped.***

**Note**: *This item is directed to measures that are risk-adjusted (with or without social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

**2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications** (*describe the steps—do not just name a method; what statistical analysis was used*)

Not applicable

**2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications?** (*e.g., correlation, rank order*)

Not applicable

**2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications?** (i.*e., what do the results mean and what are the norms for the test conducted*)

Not applicable

---

**2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS**

**2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased** due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

We ran frequencies of missing data for each self-care data element at admission and discharge as well as each of the risk adjustors after applying the exclusion criteria to examine the extent and distribution of missing data. Missing data on the IRF-PAI is identified as a dash (-), which is coded by providers to indicate they have "No information." Dash use is expected to be a rare occurrence and coding guidance is provided through in-person and web-based trainings, training manuals, and responses to help desk inquiries.

**2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data?** (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

The frequencies of missing data for each self-care data element using data from the IRF-PAI are reported in **Table 18** at admission and discharge. Across all self-care data elements, at admission and discharge, the number of cases in which the data element data are missing is very low – less than 0.1%.

**Table 18. Self-Care Data Elements: Missing Data (n=427,639)**

| | Admission: Not Assessed (-) | Discharge: Not Assessed (-) |
|---|---|---|
| **Self-Care Data Elements** | | |
| GG0130A: Self-Care - Eating | 31 (< 0.01%) | 58 (< 0.01%) |
| GG0130B: Self-Care - Oral Hygiene | 56 (<0.01%) | 56 (< 0.01%) |
| GG0130C: Self-Care - Toileting Hygiene | 65 (<0.01%) | 39 (< 0.01%) |

| | Admission: Not Assessed (-) | Discharge: Not Assessed (-) |
|---|---|---|
| GG0130E: Self-Care - Shower/Bathe Self | 88 (< 0.01%) | 62 (< 0.01%) |
| GG0130F: Self-Care - Upper Body Dressing | 39 (< 0.01%) | 42 (< 0.01%) |
| GG0130G: Self-Care - Lower Body Dressing | 28 (< 0.01%) | 36 (< 0.01%) |
| GG0130H: Self-Care - Putting on/Taking Off Footwear | 47 (< 0.01%) | 70 (< 0.01%) |
| Total | 354 (< 0.1%) | 363 (< 0.1%) |

Source: RTI analysis of IRF-PAI, January – December 2017. (Program reference: MV45).

The frequencies of missing data for each of the risk adjustors (available upon request) is also very low, ranging from no missing data for Age and Primary Diagnosis to 0.1% for the BIMS. Though missing data is rare, it is still accounted for in the calculation of the risk adjustors. For example, when determining Prior Surgery from the J2000 data element, a dash (-) on the IRF-PAI is recoded to "0" to indicate no Prior Surgery rather than dropping the patient from the performance measure calculation.

**2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased** due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

There is a very small percentage of cases with missing data, and we believe this very small percentage is unlikely to cause significant bias.

## Appendix A

**Table A-1. Intercept and Risk-Adjustor Definitions and Covariate Values for the Change in Self-Care Measure, (NQF #2633)**

| Risk Adjustor | Risk Adjustor Category | IRF-PAI Item(s) and Calculations | Intercept and Coefficients for Change in Self-Care (NQF #2633) All values have 4 decimal places |
|---|---|---|---|
| Intercept | -- | -- | 23.6369 |
| Age Group | <35 years | Truncate (Item 12 – Item 6) = age; If age <35 years = 1; else = 0 | 0.0363 |
| Age Group | 35-44 years | Truncate (Item 12 – Item 6) = age; If age 35–44 years = 1; else = 0 | 0.2755 |
| Age Group | 45-54 years | Truncate (Item 12 – Item 6) = age; If age 45–54 years = 1; else = 0 | 0.4569 |
| Age group | 55-64 years | Truncate (Item 12 – Item 6) = age; If age 55–64 years = 1; else = 0 | 0.2666 |
| Age Group | 75-84 years | Truncate (Item 12 – Item 6) = age; If age 75–84 years = 1; else = 0 | -0.5869 |
| Age Group | 85-90 years | Truncate (Item 12 – Item 6) = age; If age 85–90 years = 1; else = 0 | -1.2785 |
| Age Group | >90 years | Truncate (Item 12 – Item 6) = age; If age >90 years = 1; else = 0 | -2.0677 |

| Risk Adjustor | Risk Adjustor Category | IRF-PAI Item(s) and Calculations | Intercept and Coefficients for Change in Self-Care (NQF #2633) All values have 4 decimal places |
|---|---|---|---|
| Admission Self-Care - continuous form | Admission Self-Care - continuous form | Admission Self-Care Score = (GG0130A1 + GG0130B1 + GG0130C1 + GG0130E1 + GG0130F1 + GG0130G1 + GG0130H1) | 0.0793 |
| Admission Self-Care - squared form | Admission Self-Care - squared form | Admission Self-Care Score Squared = (GG0130A1 + GG0130B1 + GG0130C1 + GG0130E1 + GG0130F1 + GG0130G1 + GG0130H1) * (GG0130A1 + GG0130B1 + GG0130C1 + GG0130E1 + GG0130F1 + GG0130G1 + GG0130H1) | -0.0163 |
| Primary Diagnosis Group | Stroke | = 1 if Item 21A = 0001.1 or 0001.2 or 0001.3 or 0001.4 or 0001.9; else = 0 | -9.6701 |
| Primary Diagnosis Group | Non-Traumatic Brain Dysfunction | = 1 if Item 21A = 0002.1 or 0002.9; else = 0 | -5.7268 |
| Primary Diagnosis Group | Traumatic Brain Dysfunction | = 1 if Item 21A = 0002.21 or 0002.22 or 0014.1 or 0014.2; else = 0 | -4.1526 |
| Primary Diagnosis Group | Non-Traumatic Spinal Cord Dysfunction | = 1 if Item 21A = 0004.110 or 0004.111 or 0004.112 or 0004.120 or 004.1211 or 0004.1212 or 0004.130; else = 0 | -6.2803 |
| Primary Diagnosis Group | Traumatic Spinal Cord Dysfunction | = 1 if Item 21A = 0004.210 or 0004.211 or 0004.212 or 0004.220 or 004.2211 or 0004.2212 or 0004.230 or 0014.3; else = 0 | -10.1155 |
| Primary Diagnosis Group | Progressive Neurological Conditions | = 1 if Item 21A = 0003.1 or 0003.2; else = 0 | -7.0579 |
| Primary Diagnosis Group | Other Neurological Conditions | = 1 if Item 21A = 0003.3 or 0003.4 or 0003.5 or 0003.8 or 0003.9; else = 0 | -4.5214 |
| Primary Diagnosis Group | Fractures and Other Multiple Trauma | = 1 if Item 21A = 0008.11 or 0008.12 or 0008.2 or 0008.3 or 0008.4 or 0014.9; else = 0 | -5.5931 |
| Primary Diagnosis Group | Amputation | = 1 if Item 21A = 0005.1 or 0005.2 or 0005.3 or 0005.4 or 0005.5 or 0005.6 or 0005.7 or 0005.9; else = 0 | -6.9618 |
| Primary Diagnosis Group | Other Orthopedic Conditions | = 1 if Item 21A = 0006.1 or 0006.2 or 0006.9 or 0007.1 or 0007.2 or 0007.3 or 0007.9 or 0008.9; else = 0 | -5.7078 |
| Primary Diagnosis Group | Debility, Cardiorespiratory Conditions | = 1 if Item 21A = 0009 or 0010.1 or 0010.9 or 0016 or 0017.4 or 0017.51 or 0017.52; else = 0 | -4.2683 |

| Risk Adjustor | Risk Adjustor Category | IRF-PAI Item(s) and Calculations | Intercept and Coefficients for Change in Self-Care (NQF #2633) All values have 4 decimal places |
|---|---|---|---|
| Primary Diagnosis Group | Medically Complex Conditions | = 1 if Item 21A = 0011 or 0012.1 or 0012.9 or 0013 or 0015 or 0017.1 or 0017.2 or 0017.31 or 0017.32 or 0017.6 or 0017.7 or 0017.8 or 0017.9; else = 0 | -4.7541 |
| Interaction of admission self-care score and primary diagnosis group | Stroke | Admission self-care: continuous form (see above) multiplied by Primary diagnosis: Stroke (see above) | 0.2992 |
| Interaction of admission self-care score and primary diagnosis group | Non-Traumatic Brain Dysfunction | Admission self-care: continuous form (see above) multiplied by Primary diagnosis: Non-Traumatic Brain Dysfunction (see above) | 0.1598 |
| Interaction of admission self-care score and primary diagnosis group | Traumatic Brain Dysfunction | Admission self-care: continuous form (see above) multiplied by Primary diagnosis: Traumatic Brain Dysfunction (see above) | 0.0986 |
| Interaction of admission self-care score and primary diagnosis group | Non-Traumatic Spinal Cord Dysfunction | Admission self-care: continuous form (see above) multiplied by Primary diagnosis: Non-Traumatic Spinal Cord Dysfunction (see above) | 0.2025 |
| Interaction of admission self-care score and primary diagnosis group | Traumatic Spinal Cord Dysfunction | Admission self-care: continuous form (see above) multiplied by Primary diagnosis: Traumatic Spinal Cord Dysfunction (see above) | 0.3498 |
| Interaction of admission self-care score and primary diagnosis group | Progressive Neurological Conditions | Admission self-care: continuous form (see above) multiplied by Primary diagnosis: Progressive Neurological Conditions (see above) | 0.2097 |
| Interaction of admission self-care score and primary diagnosis group | Other Neurological Conditions | Admission self-care: continuous form (see above) multiplied by Primary diagnosis: Other Neurological Conditions (see above) | 0.1601 |
| Interaction of admission self-care score and primary diagnosis group | Fractures and Other Multiple Trauma | Admission self-care: continuous form (see above) multiplied by Primary diagnosis: Fractures and Other Multiple Trauma (see above) | 0.1739 |
| Interaction of admission self-care score and primary diagnosis group | Amputation | Admission self-care: continuous form (see above) multiplied by Primary diagnosis: Amputation (see above) | 0.2112 |
| Interaction of admission self-care score and primary diagnosis group | Other Orthopedic Conditions | Admission self-care: continuous form (see above) multiplied by Primary diagnosis: Other Orthopedic Conditions (see above) | 0.1773 |

| Risk Adjustor | Risk Adjustor Category | IRF-PAI Item(s) and Calculations | Intercept and Coefficients for Change in Self-Care (NQF #2633) All values have 4 decimal places |
|---|---|---|---|
| Interaction of admission self-care score and primary diagnosis group | Debility, Cardiorespiratory Conditions | Admission self-care: continuous form (see above) multiplied by Primary diagnosis: Debility, Cardiorespiratory Conditions (see above) | 0.1448 |
| Interaction of admission self-care score and primary diagnosis group | Medically Complex Conditions | Admission self-care: continuous form (see above) multiplied by Primary diagnosis: Medically Complex Conditions (see above) | 0.1533 |
| Prior surgery | Surgical | =1 if J2000 = 1; else = 0 | 0.1952 |
| Prior functioning: self-care | Dependent | =1 if GG0100A = 1; else = 0 | -4.3325 |
| Prior functioning: self-care | Some help | =1 if GG0100A = 2; else = 0 | -1.9657 |
| Prior functioning: indoor ambulation (combined) | Dependent, Some help | =1 if GG0100B = 1 or GG0100B = 2; else = 0 | -0.8074 |
| Prior Mobility Device/Aid | Walker | =1 if GG0110D = 1; else = 0 | -0.0560 |
| Prior Mobility Device/Aid | Wheelchair/Scooter Full Time/Part Time | =1 if GG0110A = 1 or GG0110B = 1; else = 0 | -0.9244 |
| Prior Mobility Device/Aid | Mechanical Lift | =1 if GG0110C =1; else = 0 | -1.9134 |
| Prior Mobility Device/Aid | Orthotics/Prosthetics | =1 if GG0110E = 1; else = 0 | -0.4816 |
| Stage 2 Pressure Ulcer | Present | =1 if M0300B1 ≥ 1; else = 0 | -0.9422 |
| Stage 3, 4 or Unstageable Pressure Ulcer | Present | =1 if M0300C1 ≥ 1 or M0300D1 ≥ 1 or M0300E1 ≥ 1 or M0300F1 ≥ 1 or M0300G1 ≥ 1; else = 0 | -1.3024 |
| Cognitive Function: Brief Interview for Mental Status score | Moderately Impaired | =1 if C0500 = 8, 9, 10, 11, or 12 or ([C0900A = 1 and C0900B = 1] or [C0900B = 1 and C0900C = 1] or [C0900A = 1 and C0900C = 1]) or [C0900A = 1 and C0900E = 1] or [C0900B = 1 and C0900E = 1] or [C0900C = 1 and C0900E = 1]); else = 0 | -0.7763 |

| Risk Adjustor | Risk Adjustor Category | IRF-PAI Item(s) and Calculations | Intercept and Coefficients for Change in Self-Care (NQF #2633) All values have 4 decimal places |
|---|---|---|---|
| Cognitive Function: Brief Interview for Mental Status score | Severely Impaired | =1 if C0500 = ≤ 7 or (C0900Z = 1 or ([C0900A=1 and C0900B = 0, and C0900C = 0, and C0900E = 0] or [C0900B=1 and C0900A = 0, and C0900C = 0, and C0900E = 0] or [C0900C=1 and C0900A = 0, and C0900B = 0, and C0900E = 0] or [C0900E=1 and C0900A = 0, and C0900B = 0, and C0900C = 0]); else = 0 | -1.7718 |
| Communication Impairment | Moderate to Severe | =1 if BB0800 = 1 or BB0800 = 2 or BB0700 = 1 or BB0700 = 2; else = 0 | -0.9358 |
| Bladder Incontinence | Less than daily, Daily, Always incontinent | =1 if H0350 = 2 or H0350 = 3 or H0350 = 4; else = 0 | -1.2459 |
| Bladder Incontinence | Urinary catheter | =1 if H0350 = 9; else = 0 | -1.1527 |
| Bowel Incontinence | Always incontinent | =1 if H0400 = 3; else = 0 | -2.3437 |
| Bowel Incontinence | Less than daily, Daily | =1 if H0400 = 1 or H0400 = 2; else = 0 | -0.8470 |
| Swallowing Ability | Modified Food Consistency | =1 if K0110B = 1; else = 0 | -0.6785 |
| Swallowing Ability | Tube/Parenteral Feeding | =1 if K0110C = 1; else = 0 | -0.7813 |
| Body Mass Index (BMI) | Low BMI | = 1 if BMI ≥ [12.0] AND ≤ [19.0]; = 0 if BMI < [12.0] OR BMI > [19.0]; = 0 if Item 25A = [0, 00, -] OR Item 26A = [-]; else = 0. Where: BMI = (([Item 26A] * 703) / Item 25A2) and the resulting value is rounded to one decimal place. | -0.3363 |
| Comorbidity Condition Group 1 | Viral and Late Effects Central Nervous System Infections (HCC4) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC4; =0 if Item 21A = 0017.1 or 0002.1 or 0002.9 or 0004.11 thru 0004.13; else = 0 | -0.2481 |
| Comorbidity Condition Group 2 | Other Infectious Diseases (HCC7) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #7; =0 if Item 21A = 0017.1; else = 0 | -0.3900 |
| Comorbidity Condition Group 3 | Metastatic Cancer and Acute Leukemia (HCC8) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #8; =0 if Item 21A = 0017.2; else = 0 | -1.1700 |
| Comorbidity Condition Group 4 | Lung and Other Severe Cancers (HCC9) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #9; =0 if Item 21A = 0017.2; else = 0 | -0.6272 |

| Risk Adjustor | Risk Adjustor Category | IRF-PAI Item(s) and Calculations | Intercept and Coefficients for Change in Self-Care (NQF #2633) All values have 4 decimal places |
|---|---|---|---|
| Comorbidity Condition Group 5 | Lymphoma and Other Cancers (HCC10) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #10; =0 if Item 21A = 0017.2; else = 0 | -0.4787 |
| Comorbidity Condition Group 6 | Diabetes: Diabetes with Chronic Complications (HCC18) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #18; =0 if Item 21A = 0017.31, 0017.32; else = 0 | -0.0736 |
| Comorbidity Condition Group 7 | Diabetes without Complication (HCC19) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #19; =0 if Item 21A = 0017.31, 0017.32; else = 0 | -0.1306 |
| Comorbidity Condition Group 8 | Protein-Calorie Malnutrition (HCC21) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #21; =0 if Item 21A = 0017.31, 0017.32; else = 0 | -0.2603 |
| Comorbidity Condition Group 9 | Other Significant Endocrine and Metabolic Disorders (HCC23) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #23; else = 0 | -0.1020 |
| Comorbidity Condition Group 10 | Intestinal Obstruction/Perforation (HCC33) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #33; else = 0 | -0.2264 |
| Comorbidity Condition Group 11 | Bone/Joint/Muscle Infections/Necrosis (HCC39) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #39; =0 if Item 21A = 0017.1, 0017.7; else = 0 | -0.5718 |
| Comorbidity Condition Group 12 | Delirium and Encephalopathy (HCC50) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #50; else = 0 | -0.4520 |
| Comorbidity Condition Group 13 | Dementia: Dementia With Complications (HCC51) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #51; =0 if Item 21A = 0002.1, 0002.9; else = 0 | -1.3119 |
| Comorbidity Condition Group 14 | Dementia Without Complications (HCC52) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #52; =0 if Item 21A = 0002.1, 0002.9; else = 0 | -1.1398 |
| Comorbidity Condition Group 15 | Nonpsychotic Organic Brain Syndromes/Conditions (HCC53) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #53; else = 0 | -0.2646 |
| Comorbidity Condition Group 16 | Reactive and Unspecified Psychosis (HCC59) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #59; else = 0 | -0.3631 |
| Comorbidity Condition Group 17 | Tetraplegia (HCC70)* | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #70; =0 if Primary Diagnosis Group = Non-traumatic spinal cord dysfunction or Traumatic spinal cord dysfunction; else = 0 | -2.1428 |

| Risk Adjustor | Risk Adjustor Category | IRF-PAI Item(s) and Calculations | Intercept and Coefficients for Change in Self-Care (NQF #2633) All values have 4 decimal places |
|---|---|---|---|
| Comorbidity Condition Group 18 | Paraplegia (HCC71) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #71; =0 if Primary Diagnosis Group = Non-traumatic spinal cord dysfunction or Traumatic spinal cord dysfunction; else = 0 | -0.8584 |
| Comorbidity Condition Group 19 | Spinal Cord Disorders/Injuries (HCC72) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #72; =0 if Primary Diagnosis Group = Non-traumatic spinal cord dysfunction or Traumatic spinal cord dysfunction; else = 0 | -0.4498 |
| Comorbidity Condition Group 20 | Amyotrophic Lateral Sclerosis and Other Motor Neuron Disease (HCC73) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #73; =0 if Item 21A = 0003.8, 0003.9; else = 0 | -1.6164 |
| Comorbidity Condition Group 21 | Cerebral Palsy (HCC74) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #74; =0 if Item 21A = 0003.5; else = 0 | -2.1701 |
| Comorbidity Condition Group 22 | Multiple Sclerosis (HCC77) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #77; =0 if Item 21A = 0003.1; else = 0 | -0.6123 |
| Comorbidity Condition Group 23 | Parkinson's and Huntington's Diseases (HCC78) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #78; =0 if Item 21A = 0003.2 or 22A, 22B or 22C = G10; else = 0 | -0.8977 |
| Comorbidity Condition Group 24 | Seizure Disorders and Convulsions (HCC79) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #79; else = 0 | -0.3754 |
| Comorbidity Condition Group 25 | Cerebral Hemorrhage (HCC99); Ischemic or Unspecified Stroke (HCC100); Cerebrovascular Atherosclerosis, Aneurysm, and Other Disease (HCC102); Hemiplegia/Other Late Effects of CVA: Hemiplegia/Hemiparesis (HCC103); Late Effects of Cerebrovascular Disease Except Paralysis (HCC105) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #99; HCC #100; HCC #102; HCC #103; HCC #105; =0 if Primary Diagnosis Group = Stroke; else = 0 | -1.1349 |

| Risk Adjustor | Risk Adjustor Category | IRF-PAI Item(s) and Calculations | Intercept and Coefficients for Change in Self-Care (NQF #2633) All values have 4 decimal places |
|---|---|---|---|
| Comorbidity Condition Group 26 | Atherosclerosis of the Extremities with Ulceration or Gangrene (HCC106) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #106; =0 if Item 21A = 0017.4; else = 0 | -0.8867 |
| Comorbidity Condition Group 27 | Aspiration, Bacterial, and Other Pneumonias: Aspiration and Specified Bacterial Pneumonias (HCC114) | =1 in Item 24 = see Crosswalk ICD-10 codes to HCC #114; =0 if Item 21A = 17.51 or 17.52; else = 0 | -0.2540 |
| Comorbidity Condition Group 28 | Legally Blind (HCC119) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #119; else = 0 | -1.5453 |
| Comorbidity Condition Group 29 | Proliferative Diabetic Retinopathy and Vitreous Hemorrhage (HCC122); Diabetic and Other Vascular Retinopathies (HCC123) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #122; HCC #123; else = 0 | -0.7617 |
| Comorbidity Condition Group 30 | Dialysis and Chronic Kidney Disease - Stage 5: Dialysis Status (HCC134); Chronic Kidney Disease, Stage 5 (HCC136) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #134; HCC #136; =0 if Item 21A = 0017.9 or 22A, 22B or 22C = N18.5; else = 0 | -0.9853 |
| Comorbidity Condition Group 31 | Chronic Kidney Disease, Severe (Stage 4) (HCC137) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #137; =0 if 22A, 22B or 22C = N18.1 or N18.2 or N18.3 or N18.4 or N18.9; else = 0 | -0.3839 |
| Comorbidity Condition Group 32 | Urinary Obstruction and Retention (HCC142) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #142; else = 0 | -0.4780 |
| Comorbidity Condition Group 33 | Chronic Ulcer of Skin, Excluding Pressure Ulcer (HCC161) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #161; =0 if Item 21A = 0017.7; else = 0 | -0.6012 |
| Comorbidity Condition Group 34 | Major Head Injury (HCC167) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #167; =0 if Primary Diagnosis Group = Traumatic Brain Dysfunction; else = 0 | -0.2234 |
| Comorbidity Condition Group 35 | Major Fracture, Except of Skull, Vertebrae, or Hip (HCC171) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #171; =0 if Item 21A = 0008.2 or 0008.4 or 0008.9 or 0014.9; else = 0 | -1.6168 |
| Comorbidity Condition Group 36 | Complication of Specified Implanted Device or Graft (HCC176) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #176; =0 if Primary Diagnosis Code = Hip and Knee Replacements; =0 if Item 21A = 0017.8; else = 0 | -0.7985 |

| Risk Adjustor | Risk Adjustor Category | IRF-PAI Item(s) and Calculations | Intercept and Coefficients for Change in Self-Care (NQF #2633) All values have 4 decimal places |
|---|---|---|---|
| Comorbidity Condition Group 37 | Amputations: Traumatic Amputations and Complications (HCC173) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #173; =0 if Primary Diagnosis Group (calculated above) = Amputation; else = 0 | -0.0492 |
| Comorbidity Condition Group 38 | Amputation Status, Lower Limb/Amputation Complications (HCC189) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #189; =0 if Primary Diagnosis Group (calculated above) = Amputation; else = 0 | -0.2744 |
| Comorbidity Condition Group 39 | Amputation Status, Upper Limb (HCC190) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #190; =0 if Primary Diagnosis Group (calculated above) = Amputation; else = 0 | -0.1697 |
| Comorbidity Condition Group 40 | Kidney Transplant Status (HCC132) | =1 if Item 24 = see Crosswalk ICD-10 codes to HCC #132; =0 if Item 21A = 0017.8 or 0017.9; else = 0 | -0.0075 |

## Appendix B:}
## Reliability and Validity Testing

### B.1 Overview of Reliability and Validity Testing

The goal of reliability testing is to ensure that items on an assessment obtain consistent results when administered or used by different clinicians. Validity testing examines whether an item or scale measures what it is intended to measure. The functional status items underwent reliability testing at the item- and scale-level in multiple types of providers in conjunction with the Post-Acute Care Payment Reform Demonstration. Item-level testing included inter-rater reliability testing within facilities and the use of videotaped standardized patients for inter-rater reliability testing across facilities/care settings. Additional testing focused on the items and scales and included internal consistency, factor analysis, and Rasch analysis. A brief summary of this testing is provided below; full reports describing the testing are available at http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/CARE-Item-Set-and-B-CARE.html

### B.2 Traditional Inter-rater Reliability Study

The reliability of the functional items was tested in a subset of 34 providers from each of the five levels of care (acute hospitals, HHAs, IRFs, LTCHs, and SNFs) distributed across 11 geographic areas. Each provider completed a duplicate CARE Item Set (admission or discharge assessment) on 15–20 patients included in the Post-Acute Care Payment Reform Demonstration (10–15 patients in the home health setting), in accordance with the guidelines and protocols.

Providers were asked to enroll a convenience sample of a set number of Medicare patients each month, representing a range of function and acuity. The overall patient sample size for each of the functional items was 450 for self-care items and 449 for mobility items (448 for transfers). After exclusions for missing data (unknown/not attempted/inapplicable), the effective sample sizes for the reliability testing were as follows:

- Eating: 401
- Oral hygiene: 414
- Toilet hygiene: 416

- Upper body dressing: 420
- Lower body dressing: 413
- Lying to sitting on the side of the bed: 412
- Sitting to standing: 387
- Chair/bed to chair transfer: 392
- Toilet transfer: 361
- Walk 150 feet: 68
- Walk once standing: 52
- Wheel in room: 46

The inter-rater reliability study included patients who were assessed by two different clinicians (raters), and the agreement of the clinicians' rating was calculated. Clinicians were instructed to have pairs of raters complete both patient assessments at the same time. Responses to items were obtained by direct observation of the patient by the clinician, and occasionally, supplemented by one or more of the following predetermined, matched methods: patient interviews (with each team member taking turns conducting and observing patient interviews); interviews with relatives/caregivers of the patient for certain items; and/or interviews with staff caring for the patient and/or chart review. Rater pairs were instructed to determine in advance which methods would be used to score the particular CARE items and to have both raters use the same methods. Raters were encouraged to divide hands-on assistance to the patient as evenly as possible for items that required hands-on assistance. Raters were instructed not to discuss item scoring during the assessment, nor to share item scores until the data were entered into the study database and finalized. Providers submitted data via the online CARE application for both assessments in each pair.

For categorical items, kappa statistics (kappa) indicate the level of agreement between raters using ordinal data, taking into account the role of chance agreement. The ranges commonly used to judge reliability based on kappa are as follows: ≤ 0 = poor; 0.01–0.20 = slight; 0.21–0.40 = fair; 0.41–0.60 = moderate; 0.61–0.80 = substantial; and 0.81–1.00 = almost perfect.

For categorical items with only two responses available, RTI International calculated only unweighted kappas. For items with more than two responses, RTI calculated both weighted and unweighted kappas. Unweighted kappa assumes the same "distance" between every one-unit difference in response across an ordinal scale. RTI used Fleiss-Cohen weights, or quadratic weights, which approximate the intra-class correlation coefficient and are commonly used for calculating weighted kappas. This choice of weighting is consistent with prior analyses of assessment reliability, where the method for developing weights was specified.7,8 Fleiss-Cohen weights put lower emphasis on disagreements between responses that fall near each other on an item scale. It should also be noted that the value of kappa can be influenced by the prevalence of the outcome or characteristic being measured. If the outcome or characteristic is rare, the kappa will be low because kappa attributes the majority of agreement among raters to chance. Kappa is also influenced by bias, and if the effective sample size is small, variation may play a role in the results. Hence, we report both weighted and unweighted kappas to give the range of agreement found under the two sets of assumptions.

Additionally, RTI calculated a separate set of kappa statistics (unweighted and weighted, where applicable) for items where additional responses outside of an ordinal scale were available (letter codes) and were set to missing.

---

[7] Hirdes JP, Smith TF, Rabinowitz T, et al. The Resident Assessment Instrument-Mental Health (RAI-MH): inter-rater reliability and convergent validity. *J Behav Health Serv Res.* 29(4):419-432, 2002

[8] Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. *Oxford University Press*, 1995.

For the traditional reliability study, kappa statistics indicated substantial agreement among raters. The weighted kappa values for the self-care items range between 0.798 for eating to 0.869 for upper-body dressing. Unweighted kappas ranged from 0.598 for oral hygiene to 0.634 for upper-body dressing. Provider-specific analyses of core self-care items show similar agreement to the overall estimates. The lower-body dressing item had the highest overall weighted kappa (0.855), whereas the eating item had the lowest (0.798). Unweighted overall kappas ranged from 0.636 (toileting) to 0.598 (oral hygiene). Acute hospitals had the highest weighted kappas across all self-care items.

The weighted kappa values for the mobility items ranged between 0.558 for walk 150 feet to 0.901 for sitting to standing and chair/bed to chair transfer. Unweighted kappas ranged from 0.667 for walk once standing to 0.762 for sit to stand. Provider-specific analyses of core mobility items show similar agreement to the overall estimates. The sit-to-stand and chair transfer items both had a weighted kappa of 0.901, whereas the lying to sitting item had a weighted kappa of 0.855. Unweighted overall kappas ranged from 0.693 (lying to sitting) to 0.762 (sitting to standing).

### B.3 Videotaped Standardized Patients Reliability Study

For the video reliability study, which was designed to examine the level of clinician agreement across care settings, clinicians in each setting were asked to assess "standardized" patients presented through a videotape of a patient assessment. This ensured that the same information was presented to each clinician and allowed examination of differences in scoring effects among different clinicians examining the "same" patient.

The patient "case studies" in each of the videos varied in terms of medical complexity, functional abilities, and cognitive impairments. The nine videos included patients classified as high, medium, or low ability/complexity for each of these three areas. Each facility or agency received three videos, one of which demonstrated one of the following elements: cognitive impairments, skin integrity problems, a wheelchair-dependent patient, and a variety of mid-level functional activities. The mid-level functional activities were considered to be the most challenging for clinicians to score and are thus of particular interest in establishing reliability. Each clinician involved in the video study watched three videos and assessed the patients according to the study guidelines and protocols. Each video was approximately 20 minutes long and had a corresponding item set arranged in the sequence in which the items appeared in the video.

The sample included 28 providers (550 assessments), which included 3 acute hospitals (15 assessments [3%]); 9 HHAs (118 assessments [22%]); 8 IRFs (237 assessments [43%]); 3 LTCHs (114 assessments [21%]); and 5 SNFs (66 assessments [12%]). Participating providers included case managers (6% of assessments), occupational therapists (14% of assessments), physical therapists (21% of assessments), registered nurses (47% of assessments), speech therapists (5% of assessments), and others, mostly licensed practical nurses (LPNs; 8% of assessments).

Two main analytic approaches were used for assessing the video reliability of the CARE items, adhering closely to the methods used by Fricke et al.[9] in their video reliability study of the FIM®[10] instrument. First, percent agreement with the mode response was calculated for each CARE item included in at least one of the nine videos. Unlike the approach used by Fricke et al., RTI did not consider agreement at one response level above and below the mode, and instead used a stricter approach looking at direct modal agreement only. In the second approach, percent agreement with the internal clinical team's consensus response was also calculated. This second measure not only gives an indication of item reliability, but also reflects training consistency for the providers.

The video reliability study indicated substantial agreement with the mode and clinical team among all items, typically upwards of 70%. The notable exception to this trend exists among the clinicians in the "Other"

---

[9] Fricke J, Unsworth C, Worrell D. Reliability of the Functional Independence Measure with Occupational Therapists. *Australian Occupational Therapy Journal* 40(1):7-15, 1993.

[10] FIM® is a trademark of Uniform Data System for Medical Rehabilitation, a division of UB Foundation Activities, Inc.

category (mostly LPNs); they consistently had the lowest levels of agreement among all core self-care items, ranging from 50 to 72%. For the toileting and dressing items, the agreement with the clinical team was lower than with the mode. This occurred because the clinical team response differed from the mode for these three items in either one or two videos. Nonetheless, because the clinical team response and mode were identical on most of the videos, agreement was still quite high for these items. In general, study clinicians had responses on average that agreed with the expert clinical team or were slightly lower.

The video reliability study indicated substantial agreement with the mode and clinical team for the lying-to-sitting, sit-to-stand, chair/bed to chair transfer, and toilet transfer items (greater than 76%). Although rates of agreement with the mode and clinical team response were generally identical, for the toilet transfer item, the clinical team agreement is slightly lower. The items for walking and wheeling distances showed more variable levels of agreement across disciplines, with overall agreement generally in the moderate range (50–78%). For the Walk in Room item, there was a notable decrease in the agreement with the clinical team compared to agreement with the mode. This occurred because in two of the four videos where this item was assessed, the clinical team response differed from the mode.

### B.4 Scale-level Reliability Results: Internal Consistency

In addition to item-level reliability testing, we examined internal consistency, which provides a general assessment of how well the items interrelate within a domain or subscale. Internal consistency is assessed using the Cronbach's alpha coefficient, which is the average correlation of all possible half-scale divisions. Cronbach's alpha is a statistic frequently assessed when instrument or scale psychometrics are published. The Cronbach's alpha reliability estimate ranges from zero to one, with an estimate of zero indicating that there is no consistency of measurement among the items, and one indicating perfect consistency. Many cutoff criteria exist to determine whether or not a scale shows good consistency or whether the items "hang together" well. General consensus is that Cronbach's alpha should be at least 0.70 for an adequate scale for group-level decisions, and alphas closer to 1 indicate a good scale.11

Assessments of individual self-care and mobility subscales at both admission and discharge tend to show good reliability statistics (Cronbach's Alpha of at least 0.80) within their specified subscales. Reliability estimates by provider type show that the functional status items maintain a very high internal consistency. In addition, no one provider type appears to have reliability estimates higher or lower than the rest, indicating similarity of CARE usage with respect to internal consistency.

The following table shows the findings from the Cronbach's alpha internal consistency evaluation mentioned above.

**Table B-1**
**CARE functional status internal consistency reliability summary by provider type**

| CARE analytic set | Overall alpha | HHA alpha | SNF alpha | IRF alpha | LTCH alpha |
|---|---|---|---|---|---|
| Self-Care | 0.96 | 0.94 | 0.95 | 0.95 | 0.96 |
| Mobility | 0.96 | 0.94 | 0.95 | 0.96 | 0.97 |

### B.5 Scale-level Reliability and Validity Testing: Rasch Analysis

Because we are measuring a latent trait—a concept that is not measured directly, but that relies on activities that can be directly observed—we used the one-parameter Rasch model to gain a better understanding of the functional status activities. More specifically, we examined the order of functional status items (from least challenging to most challenging) that characterize the concepts of the self-care and mobility.

Rasch analysis uses the scores from the functional assessment items to create the equivalent of a functional status "ruler" (i.e., scale). Rasch analysis uses the available data to estimate a person's location along the

[11] Aron A, Aron EN *Statistics for Psychology*. 2nd ed. Upper Saddle River, NJ: Prentice Hall, 1999.

"ruler;" therefore, analyses can be conducted if some data are missing. Rasch analysis can also inform the optimal selection of key items in order to construct functional status scales that sufficiently span an entire range of patient functioning, so that both the least able and most able (lowest- and highest-functioning) patients are adequately measured. In addition, Rasch analysis can indicate where items overlap or are redundant in terms of the level of function they capture.

Rasch analysis has been used to examine the FIM® instrument,[12,13,14,15] the Minimum Data Set (MDS),[16] and the Outcome and Assessment Information Set (OASIS).[17] Rasch analysis has also been used to examine the extent to which existing functional assessment instruments (e.g., the FIM® instrument, MDS 2.0) capture the same construct.[18]

Rasch measurement is based on a probabilistic model that describes the association between a person's underlying ability level and probability of a particular item response, and summarizes a patient's position along a "ruler" that represents a latent trait or concept (e.g., self-care or mobility).[19] In essence, the Rasch analysis creates a ruler based on the domain measured (e.g., mobility) that can be used to assess the abilities of the patients. The analysis also provides information on the hierarchy of item difficulty (from easy to hard) that can be used to evaluate the construct validity of a set of items. In addition, the Rasch analysis provides information about the level of challenge associated with each item rating scale ("dependent" through "independent"). For example, an item with a low difficulty estimate (e.g., eating) would be more likely to be completed with little or no help by patient's items that are more challenging (e.g., 12 steps), where most patients would find completing this activity challenging. Finally, the Rasch analysis can provide information on items that do not fit into the single theorized concept through "item misfit" statistics, which may indicate that the item needs further evaluation before it is included on future administrations of the subscale. The infit mean square is an indicator of the degree to which patient responses are similar to what would be expected (i.e., predicted) by the measurement model. The acceptable range is generally 0.6 to 1.4. If the item values are above this range, it reflects that person response patterns are erratic, generally suggesting that the item is not measuring the same construct as other items. Infit mean squares above 1.4 are considered to be

---

[12] Granger CV, Hamilton BB, Linacre JM, et al. Performance profiles of the functional independence measure. *Am J Phys Med Rehabil.* 72(2):84-89, 1993.

[13] Linacre JM, Heinemann AW, Wright BD, et al. The structure and stability of the Functional Independence Measure. Archives of Physical Medicine & Rehabilitation.75(2):127-132, 1994

[14] Wright BD, Linacre JM, Smith RM, et al. FIM measurement properties and Rasch model details. Scandinavian Journal of Rehabilitation Medicine, 29(4):267-272, Dec. 1997.

[15] Heinemann AW, Linacre JM, Wright BD, et al. Relationships between impairment and physical disability as measured by the functional independence measure. Arch Phys Med Rehabil. 74(6):566-573, 1993.

[16] Wang YC, Byers KL, Velozo CA. Rasch analysis of Minimum Data Set mandated in skilled nursing facilities. J Rehabil Res Dev. 45(9):1385-1399, 2008.

[17] Fortinsky RH, Garcia RI, Joseph Sheehan T, et al. Measuring disability in Medicare home care patients: application of Rasch modeling to the outcome and assessment information set. Med Care. 41(5):601-615, 2001.

[18] Velozo CA, Byers KL, Wang YC, et al. Translating measures across the continuum of care: using Rasch analysis to create a crosswalk between the Functional Independence Measure and the Minimum Data Set. J Rehabil Res Dev. 44(3):467-478, 2007.

[19] Wright BD, Stone MH. Best Test Design. Rasch Measurement. 1979.

unacceptably unexpected[20] and indicate that the item most likely does not reflect the same construct as the other items included in the scale; for example, a need for assistance with self-care.

RTI used Rasch analysis to examine the extent to which the items worked together to define a coherent concept. This was conducted separately for the self-care and mobility items. Item fit statistics were examined as an indication of how well all items work together to describe the overall construct (self-care or mobility). The Rasch analysis provides insight into how the items work together as a subscale, including the hierarchy of item difficulty (ordering from easy to difficult) and item fit to the model.

Examinations of these Rasch analysis results reveal that the mobility and self-care item hierarchies make sense clinically and that the operational definitions of the constructs maintain general stability from admission to discharge. Some items have fit statistics outside the acceptable range (e.g., pick up object from floor), but members of the Technical Expert Panel noted that this is an important assessment given the risk of falls.

RTI examined how well the items selected measure the persons in the data set for both self-care and mobility items. RTI examined the extent to which person response patterns fit the assumptions of the measurement model using the same range of infit statistics identified above. RTI examined the extent to which persons are effectively measured (ceiling and floor effects) in each setting overall and for admission and discharge time points. The mobility and self-care items were found to be well targeted to the range of patient ability sampled within this post-acute care population.

RTI established that the six steps of the CARE rating scale are operating as intended, both overall and for individual items on the self-care and mobility subscales. The probability that a person will be scored on a particular rating scale step varies depending on the functional ability of the person. That is, very able people will be more likely to be scored as '5' and '6' than as '1' and '2.' Looking empirically at these distributions, one should see the transitions from one step to the next (called thresholds) proceed monotonically and distinctly across the range of person abilities. In other words, there should always be some point along the range at which each rating-scale step is more probable than another step. When a rating-scale step is not more probable at any point, it suggests that raters are not able to use that step to consistently distinguish patient ability at that level.

## 3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

**3a. Byproduct of Care Processes**

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

**3a.1. Data Elements Generated as Byproduct of Care Processes.**

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

If other:

**3b. Electronic Sources**

[20] Wright BD, Linacre JM, Gustafson J, et al. Reasonable mean-square fit values. Rasch Measurement Transactions. 8(3):370, 1994.

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

**3b.1. To what extent are the specified data elements available electronically in defined fields** (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.

ALL data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)

**3b.2. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.** For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

Not applicable. This quality measure's data elements are collected solely from electronic sources.

**3b.3. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.**

**Attachment:**

**3c. Data Collection Strategy**

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

**3c.1. Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

**IF instrument-based, consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.**

The NQF feasibility criterion requires measure developers to: 1) demonstrate that the data collection strategy can be implemented and 2) describe any difficulties regarding data collection.

Data Collection:

Data for this quality measure are currently collected and submitted to the Centers for Medicare and Medicaid Services using the Inpatient Rehabilitation Facility Patient Assessment Instrument (IRF-PAI). These data have been collected by all IRFs in the US since October 1, 2016 as part of the IRF Quality Reporting Program (QRP). In addition, beginning in October 2019, data from Section GG will also be required by the Centers for Medicare and Medicaid Services as part of the IRF Prospective Payment System.

The measure data are "generated" by qualified clinicians as they observe patients completing daily activities, such as eating and oral hygiene at the time of admission and discharge. As shown in the testing form, missing data is minimal (less than 0.1% across all data elements). The IRF-PAI data are submitted to CMS via the QIES ASAP system, which has been in place since 2002. This data submission system is secure and encrypted with administrative, physical and technical safeguards in place.

Preventing and Addressing Potential Data Collection Challenges:

The Centers for Medicare and Medicaid Services finalized the implementation of this quality measure in August 2014 in the FY 2015 IRF PPS Final Rule, more than 1 year before implementation of data collection. This advance notice allowed providers, vendors and CMS to prepare for implementation. The Centers for Medicare

and Medicaid has developed software that is free for IRF to use to submit IRF-PAI data. Also, given the Centers for Medicare and Medicaid's many years of experience with data submission (the IRF-PAI data have been submitted to the Centers for Medicare and Medicaid since 2002) implementation occurred with minimal difficulty.

To assist providers with the collection of accurate data, the Centers for Medicare and Medicaid Services has offered multiple in-person and on-line training opportunities since May 2015. In addition, a help desk is available to answer provider questions regarding data collection, and "Q & A" documents are posted on the CMS website for provider use. Training information is available on the following website: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/IRF-Quality-Reporting/IRF-Quality-Reporting-Training.html

**3c.2. Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.

There no costs associated with fees, licensing or other requirements associated with the measure data elements or risk model.

## 4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

**4a. Accountability and Transparency**

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

**4.1. Current and Planned Use**

*NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.*

| Specific Plan for Use | Current Use (for current use provide URL) |
|---|---|
| | Public Reporting<br>https://www.medicare.gov/inpatientrehabilitationfacilitycompare/<br>Measure data from calendar year 2019 (currently being collected) will be publicly reported on IRF Compare in 2020 for the IRF Quality Reporting Program<br>Quality Improvement (external benchmarking to organizations)<br>IRF QRP: On confidential feedback reports and IRF Compare, providers can view national-level performance measure scores for benchmarking quality efforts. IRFs can also review and compare scores for local providers through IRF Compare's web features.<br>https://qtso.cms.gov/<br>Quality Improvement (Internal to the specific organization)<br>IRF QRP: IRFs receive confidential feedback reports through the CMS designated data submission system, which includes the Review and Correct, Quality Measure, and Provider Preview Reports to review their data internally.<br>https://qtso.cms.gov/ |

**4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

Name of Program and Sponsor and Purpose:
This quality measure has been implemented in the Center for Medicare and Medicaid's (CMS) Inpatient Rehabilitation Facility Quality Reported Program (IRF QRP) and serves two purposes:
1) to share quality data with each IRF that may be used to support quality improvement efforts; and
2) to share quality data about each IRF, which may assist consumers and family members in making decisions about where to receive IRF care.
As part of the IRF QRP, IRFs have been able to view data for this quality measure in their confidential feedback reports, which may be used for quality improvement, since April 2017.
Quality measure data collected in calendar year 2019 will be publicly reported in 2020 on CMS's IRF Compare website at: https://www.medicare.gov/inpatientrehabilitationfacilitycompare/. Since 2016, CMS has publicly reported IRF QRP quality measure data on the IRF Compare website. This website reports quality data for each IRF, and these data are also publicly available for download at: https://data.medicare.gov/data/inpatient-rehabilitation-facility-compare.
This measure was implemented pursuant to two public laws that addressed the IRF QRP and reporting of data submitted by providers:

1) The Patient Protection and Affordable Care Act ("ACA") of 2010 (Public Law No: 111-148)
    - Section 3004(b) of the ACA amended section 1886(j)(7) of the Social Security Act (SSA) requiring the Secretary to establish quality reporting requirements for IRF providers. Quality reporting applies to all IRF providers receiving payment under the IRF Prospective Payment System (PPS).
    - The ACA mandates IRFs to submit data or be subject to a two-percent reduction in their annual payment update (APU) determination.
2) The Improving Medicare Post-Acute Care Transformation Act ("IMPACT Act") of 2014 (Public Law No: 113-185):
    - The IMPACT Act requires IRFs to submit standardized patient assessment data on quality, resource use, and other measures.
    - The data submitted from providers are used to calculate measures that report healthcare processes and patient outcomes among IRF providers under the QRP.
    - Requires the establishment of procedures for making provider performance information available to the public.

CMS finalized in the FY 2019 IRF PPS final rule (83 FR 38562) that they plan to publicly report data for this performance measure on IRF Compare in the fall of 2020. The first time the data will be publicly displayed will be for patients discharged on January 1, 2019 through December 31, 2019.

CMS provides an opportunity for IRFs to review their own data before it is publicly reported through confidential feedback reports available in the CMS designated data submission system. Several reports are available that provide different snapshots of the measure data (described in more detail below in 4a2.1.1). As of April 2017, providers could view the observed change in self-care performance measure in their confidential Review and Correct reports. The risk-adjusted change in self-care performance measure became available in the Quality Measure reports October 2018.

Geographic Area, Accountable Entities and Patients Included:

The IRF QRP measures are calculated for 100% of IRF providers in the US (1,129 IRFs in 2017). This includes IRFs in every US state, the District of Columbia, and the US Territory of Puerto Rico. IRFs submitted a total of 493,209 IRF-PAI records for Medicare Part A and Medicare Advantage patients discharged in 2017.

All providers receive their confidential feedback reports, which may be used for internal quality improvement efforts.

To ensure reliability of the performance measure scores, IRFs with less than 20 patients (12 IRFs in 2017) during a reporting period would not have their data displayed publicly. Once an IRF has more than 20 patients during the reporting period, their data would display on IRF Compare.

Level of Measurement and Setting:

As mentioned, this quality measure has been implemented in the IRF setting as part of the IRF QRP. The measure score is reported at the facility-level.

**4a1.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

Not applicable because public reporting is currently underway for this measure.

**4a1.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

Not applicable because public reporting is currently underway for this measure.

**4a2.1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.**

**How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.**

For Providers:

Dissemination of performance results and assistance with interpretations of the performance data for IRFs have been addressed in four specific ways: confidential feedback reports, provider training seminars, manuals and materials, responses to questions submitted to the IRF QRP Help Desk: IRF.Questions@cms.hhs.gov, and IRF Public Reporting Help Desk: IRFPRQuestions@cms.hhs.gov, and on IRF Compare.

1) Confidential Provider Feedback Reports:

All IRFs who submit IRF-PAI data to CMS receive three types of confidential reports with performance measure data and scores based on the data submitted. These reports support internal quality improvement efforts and include the Review and Correct, Quality Measure, and Provider Preview Reports. Details about each of these reports is provided below in 4a.2.1.2.

2) IRF QRP Provider Training Seminars:

CMS conducted several in-person IRF QRP provider training seminars to share information about coding the data elements used to calculate the performance measure, to share details about the measure specifications and to explain how the measure is calculated. Training sessions that focused on the confidential feedback reports were also conducted to support providers in reviewing and interpreting the data they receive in these reports. During training sessions, providers were encouraged to ask questions about coding the data elements and the change in self-care performance measure to ensure an accurate understanding of the measure. Training materials are posted on the CMS website after each training seminar is completed. To review provider training materials, see the following webpage:

https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/IRF-Quality-Reporting/IRF-Quality-Reporting-Training.html

The IRF QRP Measure Calculations and Reporting User's manual, which presents the measure specifications and how the measures are calculated for each measure in the IRF QRP, is posted on the CMS website.

Therefore, providers have detailed measure specifications available to them. To review the current IRF QRP Measure Calculations and Reporting User's manual, see the following webpage:

https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/IRF-Quality-Reporting/Downloads/IRF-Measure-Calculations-and-Reporting-Users-Manual-V30.pdf

3) IRF QRP and IRF Public Reporting Help Desk:

CMS also maintains a provider help desk for the IRF QRP where IRFs can submit questions about the data elements, the measure, including questions about performance data, interpretation of results, or instructions on coding (IRF.Questions@cms.hhs.gov). A help desk for questions about the data available on IRF Compare (see below) is also available (IRFPRQuestions@cms.hhs.gov). A response is provided to address each question that is submitted.

4) IRF Compare Website:

The performance measure data are publicly displayed on the IRF Compare website and plain language is used to assist users in interpreting the data that are presented. The quality of care that IRF providers deliver to patients can vary from facility to facility, and publicly displaying performance data on IRF Compare supplies information for providers to use for improving the quality of care they provide to patients.

The IRF QRP Measure Calculations and Reporting User's manual, which presents the measure specifications and how the measures are calculated for each measure in the IRF QRP, is posted on the CMS website. Therefore, providers have detailed measure specifications available to them. To review the current IRF QRP Measure Calculations and Reporting User's manual, see the following webpage:

https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/IRF-Quality-Reporting/Downloads/IRF-Measure-Calculations-and-Reporting-Users-Manual-V30.pdf

For Patients, Families, Carers and Other Stakeholders:

IRF patients, family members, carers, and other stakeholders (researchers, journalists, policymakers) can view an IRF's measure performance information on the publicly available IRF Compare website. The IRF Compare website is designed to help patients and caregivers make informed decisions about their health care and to compare inpatient rehabilitation facilities based on important indicators of quality. Preparations to include the performance data for this measure on the IRF Compare Website includes developing plain language to explain the measure and the results for the general public. Additionally, the IRF Compare Website has gone through consumer testing to test functionality and usability. IRF Compare is available in both English and Spanish.

Furthermore, the public can download the IRF Compare datasets. The files contain general information about providers, provider level data on quality measures, and national data shown on the site. A data dictionary provides detailed information on the measures and file layouts.

Public access to the performance data on the IRF Compare website has been widespread and increasing over time. In Quarter 4 of 2017, there were over 14,000 sessions and 40% of those were returning visitors. Subsequently, the number of sessions increased by 27.6% a year later to over 18,000 sessions in Quarter 4 of 2018 in which 42% of those were returning visitors.

**4a2.1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.**

All IRFs receive three types of confidential reports with performance measure data and scores based on the data submitted:

1) Quality Measure Reports:

The intent of this report is to enable IRFs to track their own quality measure data at the facility- and patient-level. Data for this report is refreshed monthly and displays performance measure information at the facility- and patient-stay level for review. The facility-level report displays the measure denominator, average observed scores, average risk-adjusted score, and the national average for benchmarking the facility's performance. The

patient-level report displays which patients are excluded from the measure as well as each patient's observed change in self-care score.

2) Review and Correct Reports:

The intent of this report is for IRFs to view their data prior to the quarterly data submission deadline to ensure accuracy of the data submitted to CMS. Data for this report is refreshed weekly and displays data correction deadlines and whether the data correction period is open or closed. Only the last four quarters of data are available in this report.

3)Provider Preview Reports:

The intent of this report is for IRFs to preview what performance data will publicly displayed for their IRF. The report displays facility-level performance measure data and shows risk-adjusted values and national rates as they will appear publicly on IRF Compare. Data displayed in this report cannot be modified by the provider.

**4a2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.**

**Describe how feedback was obtained.**

In addition to the processes and information described above in 4a2.1.1 and 4a2.1.2, CMS solicited public comments about the change in self-care performance measure via a 60-day public comment period during the fiscal year (FY) 2016 rulemaking process. CMS also solicited public comments during the FY 2019 rulemaking process on the proposal to publicly report this measure on IRF Compare.

See below for links to the final rules which present all public comments received and responses:

FY 2016: https://www.federalregister.gov/articles/2015/08/06/2015-18973/medicare-program-inpatient-rehabilitation-facility-prospective-payment-system-for-federal-fiscal

FY 2019: https://www.federalregister.gov/documents/2018/08/06/2018-16517/medicare-program-inpatient-rehabilitation-facility-prospective-payment-system-for-federal-fiscal

**4a2.2.2. Summarize the feedback obtained from those being measured.**

We received support for both implementation and public reporting of the change in self-care performance measure for the IRF QRP. Comments were received from various stakeholders, including providers, provider associations, researchers, government agencies, information system vendors, advocacy groups, and individuals/consumers.

In the FY 2016 rule proposal, most commenters supported the change in self-care performance measure being added to the IRF QRP and stated that this measure contributes to meaningful differences in IRF patients' outcomes. Several commenters supported the risk adjustment model, specifically highlighting the inclusion of prior mobility device use and prior functioning as important risk adjustors for functional outcome measures. Commenters encouraged CMS to continue to examine data for this quality measure and to improve the risk adjustment methodology over time. Several commenters requested that CMS provide additional reliability and validity testing and recommended training programs to ensure data accuracy.

In the FY 2019 rule proposal, most commenters supported publicly reporting this measure. Some provided recommendations on how to publicly display the measure, including a consumer-friendly name and adequate consumer testing to develop appropriate language for explaining the measure to the public. Concerns were noted about publicly reporting the measure before providers have enough time to review their data, track their performance and ensure that their provider-level performance is accurately represented on IRF Compare.

Additional feedback by providers is also regularly received through the active IRF QRP help desk. As noted above, IRF staff submit questions about the measure, including questions about performance data, interpretation of results, or instructions on coding to the IRF QPR help desk. Individuals viewing the measure data on IRF Compare can submit questions or comments to the IRF Public Reporting help desk. Through these avenues, CMS receives ongoing, real-time feedback which further supports measure improvement and maintenance.

As part of CMS's ongoing efforts to engage stakeholders in the measure development, improvement and refinement process, all comments and questions are taken into consideration. Several points of feedback were tested and are planned for future measure implementation (see 4a2.3 below for examples).

**4a2.2.3. Summarize the feedback obtained from other users**

In March 2017, the measure developer convened stakeholders and experts who contributed direction and thoughtful input for IRF QRP measure development and maintenance. This technical expert panel (TEP) was asked to discuss and make future recommendation on the change in self-care performance measure. Feedback included general support for the outcome measure and suggestions for new risk adjustors. The TEP noted that plain language descriptions of the measures would be important to assist consumers' ability to interpret the function change scores when posted on IRF Compare.

The IRF QRP TEP Summary report is available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/IRF-Quality-Reporting/Downloads/2017-IRF-QRP-TEP-Summary-Report-_508C.pdf

Additional feedback by consumers and researchers is also received through the IRF Public Reporting help desk. Individuals viewing the IRF Compare website can submit questions or comments and, in this way, CMS provides real-time support to patients, families and carers and other stakeholders seeking additional information or clarification on measures. Researchers and academics needing assistance in understanding and using the downloadable data also submit questions. These questions and comments are used to support CMS's goal of continuously improving the website.

**4a2.3. Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.**

Part of our measure maintenance process includes incorporating stakeholder feedback as we continue examination and refinement of performance measures. CMS and RTI International reviewed and took into consideration all public comments received in both the FY 2016 and FY 2019 final rules as well as feedback from the March 2017 technical expert panel and comments and questions received via the help desks.

Updates were made to the change in self-care performance measure from the initial NQF endorsement, and these updates are partly based on stakeholder feedback. For example, commenters encouraged CMS to continue reviewing the data and improving the risk adjustment model over time which we have done for this latest measure update.

Stakeholder comments on the public display of the measure on IRF Compare were also taken into consideration. This included feedback from rulemaking public comments, the 2017 IRF TEP, and consumers. For example, consumer testing is done prior to public reporting and plain language is displayed on the website (e.g., a consumer-friendly name rather than the technical measure name). Additionally, to address industry concerns that providers needed adequate time to understand their measure data before it was publicly reported, the first data to display on IRF Compare will be calendar year 2019 (January – December 2019) though data collection began October of 2016.

**Improvement**

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

**4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)**

**If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**

The change in self-care performance measure was recently implemented on October 1, 2016 and will be publicly reported for the first time in the fall of 2020 using calendar year 2019 data. Thus, there is no extensive data to evaluate trends in performance over time. In Section 1b, we provide analysis comparing fiscal year 2017 and calendar year 2017 as well as data by quarter and show that the measure remained stable over this period. As more data becomes available, we will examine score distribution and change in provider performance scores.

**4b2. Unintended Consequences**

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

**4b2.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.**

No unexpected findings have been identified during implementation and testing of this measure. To date, no unintended impacts on patients have been identified.

**4b2.2. Please explain any unexpected benefits from implementation of this measure.**

To date, no unintended impacts on patients have been identified.

## 5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

**5. Relation to Other NQF-endorsed Measures**

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

Yes

**5.1a. List of related or competing measures (selected from NQF-endorsed measures)**

0174 : Improvement in bathing

0175 : Improvement in bed transferring

0426 : Functional status change for patients with Shoulder impairments

0427 : Functional status change for patients with elbow, wrist and hand impairments

0428 : Functional status change for patients with General orthopaedic impairments

0688 : Percent of Residents Whose Need for Help with Activities of Daily Living Has Increased (long stay)

2286 : Functional Change: Change in Self Care Score

2287 : Functional Change: Change in Motor Score

2613 : CARE: Improvement in Self Care

2643 : Average change in functional status following lumbar spine fusion surgery

2769 : Functional Change: Change in Self Care Score for Skilled Nursing Facilities

2775 : Functional Change: Change in Motor Score for Skilled Nursing Facilities

2776 : Functional Change: Change in Motor Score in Long Term Acute Care Facilities

2777 : Functional Change: Change in Self Care Score for Long Term Acute Care Facilities

**5.1b. If related or competing measures are not NQF endorsed please indicate measure title and steward.**

Change in Daily Activity Function as Measured by the AM-PAC (CREcare)

**5a. Harmonization of Related Measures**

The measure specifications are harmonized with related measures;

**OR**

The differences in specifications are justified

**5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**

**Are the measure specifications harmonized to the extent possible?**

No

**5a.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**

All the listed measures address the same topic, function, but the target populations for most of these measures is not the IRF patient population. For example, measures are used for patients/residents treated in outpatient settings, home care, skilled nursing facilities, long-stay nursing homes, and long-term care hospitals. One measure has been previously identified by NQF staff as a competing measure: Functional Change: Change in Self Care Score (NQF #2286).

**5b. Competing Measures**

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);

**OR**

Multiple measures are justified.

**5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**

**Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)**

The NQF and the Patient Experience and Function Standing Committee may choose to endorse both competing measures, because both provide value. If NQF and the committee believe that only one measure should be endorsed as "best-in-class," we offer a list of the strengths of our measure below as well as a comparison of feasibility, usability and use for consideration.

Specifically, we describe the similarities and important differences between this change in self-care measure and the listed related and competing measures (See 5.1.a). We note that several features of this measure (e.g., the data elements, many of the risk adjustors, and the risk-adjustment approach) are the same as or aligned with the specifications of several of the other endorsed measures. Therefore, we believe that the specifications for this measure incorporate the best features of all endorsed related and competing measures, and, as a whole, represents the "best in class" for measuring change in self-care for IRFs.

This Change in Self-Care (NQF #2633) measure was developed by building on the most recent science related to measurement of patient functioning and quality measure development. The latest science and scholarly literature, clinical thinking, and expert input on functional assessment and quality measurement was combined with a cross-setting design and purpose in mind. Specifications were discussed with stakeholders and experts, pilot tested, and analyzed throughout the development process, as described in the Testing form.

Functional Assessment Data Elements

1. Cross-Setting Design

The functional assessment data elements for this measure, included in Section GG: Functional Abilities and Goals, were designed and tested with a cross-setting purpose in mind to ensure that data may be collected by clinicians in various post-acute and acute care settings. This enhances the cross-setting validity and reliability of quality measures that use these data. Standardization of self-care and mobility data elements across post-acute care settings has been an important goal for policymakers and included in the IMPACT Act of 2014. We

note that another measure focused on improvement in self-care, Related Measure NQF #2613, also use the data elements from Section GG: Functional Abilities and Goals as part of their performance measure with the rationale that the data elements were developed for cross-setting use and that the data elements are standardized.

2. Clinician Observation

To determine a patient's functional ability, providers are instructed to code the data elements in Section GG: Functional Abilities and Goals primarily based on clinical observation. Specifically, a qualified clinician will assess the patient's performance based on direct observation, as well as gather input from reports from other clinicians, care staff, or family as well as the patient's self-report. Typically, an interdisciplinary team of qualified clinicians is involved in assessing the patient and CMS provides guidance through manuals, training programs, and help desk responses to support providers in collecting accurate functional assessment data. We note that the Competing Measure NQF #2286 and Related Measures NQF #2613, #2769, #2775, #2776, and #2777 also use clinician observation to assess and code a patient's functional abilities.

3. Functional Assessment Data Elements Capture A Range of Functioning

The functional assessment data elements and associated rating scale were designed to build on the existing science of functional assessment, which included a review of the strengths and limitations of existing instruments. The inclusion of 7 self-care data elements allows for the measurement of a wide range of patient functioning and thus the opportunity to demonstrate gains in a variety of functional activities. Patients may be expected to make varying amounts of improvement, from minimal to large improvement, across different activities. We note that the Related Measure NQF #2613 also use these self-care data elements to measure improvement in self-care for the Skilled Nursing Facility setting.

4. Simplified and Targeted Rating Scale

The function data elements used in this performance measure are coded using a 6-level rating scale that indicates the patient's level of independence performing an activity; higher scores indicate more independence. The decision to use a 6-level rating scale was based on several factors. First, input from the clinical communities and research examining the relationship between minutes of assistance and functional assessment scores,. Second, scores do not decrease due to the use of an assistive device, which is consistent with the approach used by the World Health Organization's International Classification of Functioning (ICF) that suggests what matters most is someone's capacity to do an activity regardless of the use of assistive devices. Thus, the 6-level rating scale was designed to measure a person's ability to perform daily activities with or without assistive devices. The rating scale focused solely on the type and amount of human assistance needed to compete an activity. Another measure of self-care function, Related Measure NQF #2613, used in the Skilled Nursing Facility, also adopted the 6-level rating scale.

5. Meaningful Activity Not Attempted Codes

The use of four distinct activity not attempted codes were implemented so that providers code a specific reason for an activity not being attempted. For example, code 07 is used if the patient refused to attempt the self-care activity, such as putting on/taking off footwear, during the entire 3-day assessment period. If the patient was not able to perform the activity safely, due to medical or safety concerns, code 88 is used. A qualified clinician's assessment that a patient's medical condition contributes to their inability to safely put on and take off footwear means something different than a patient who is refusing to perform the activity, and the coding responses that allow for this distinction. Other measures of self-care function, such as Related Measure NQF #2613 used in the Skilled Nursing Facility also adopted the activity not attempted codes.

Measure Calculation

1. Difference Approach for Interpretability

This measure calculates the risk-adjusted performance score using observed and expected scores. When observed and expected scores are compared, the difference between the two scores is calculated, and this difference approach represents an additive relationship (i.e., the observed change in function minus the

expected changed in function, plus the national average). The choice between using a difference or a ratio approach depends on the researcher's assumption on whether the relationship between risk factors and the outcome is additive or multiplicative (Mukamel et al., 2000). After we conducted testing using the two approaches, and consulted with methodological experts, we decided to use the difference approach for this measure. When the expected value is small, the ratio is more volatile with small changes in the observed values (Ash et al, 2012). As the denominator approaches zero, the ratio can increase greatly in magnitude, as the observed values become greater than the expected values. Also, if the average expected value is 0, then the ratio cannot be calculated. The following measures also use this approach: Related Measure NQF #2613, used in the Skilled Nursing Facility, and the FOTO measures (NQF# 0426, 0427, and 0428).

2. Exclusion Criteria to Maintain Validity

We believe exclusion criteria are important specifications that support the validity of the quality measure. The exclusion criteria were selected with input from the Technical Expert Panel and input from a public comment process, as well as a review of existing literature. Patients with limited or less predictable self-care due to the nature of their medical condition improvement (e.g., severe brain damage) were recommended for exclusion by experts. Their reasoning was that attributing limited improvement in patients with these conditions to poor quality of care by the IRF would threaten the validity of the quality measure. The Related Measures NQF #2613 and #0688 also exclude patients with selected medical conditions where improvement is very unlikely in order to maintain the validity of the measures' performance scores.

The measure also has exclusions for patients with incomplete stays (e.g., discharged to acute care) or patients who were discharge to hospice for whom functional improvement may not be a goal. The Related Measures NQF #2613 and #0688 also exclude hospice patients from their performance measures.

3. Robust Risk Adjustment Model

Improvement in functional abilities for patients in IRFs are associated with many patient demographic and clinical characteristics. Existing literature, stakeholder comments and technical expert opinions about risk adjustors were gathered and we all suggestions were tested with data. This measure adjusts for patient demographic and clinical characteristics, including age category, primary rehabilitation diagnosis, prior functioning, admission self-care or mobility functional status, cognitive function, communication function, and comorbidities. Adequate risk adjustment is critical to ensure quality measure validity, such that differences in performance scores across IRFs are related to differences in quality of care as much as possible, rather than to differences in patient characteristics across facilities.

For an individual patient, up to 61 risk adjustors may apply in the self-care model. Notably, 40 of these are for comorbidities. This number of comorbidities are included in the model to account for differences in functional improvement for people with different co-existing health conditions. We would like to highlight that no patient in the national data had all 40 comorbidities and, in fact, the maximum number of comorbidities a person had was 11. On average, patients had only 1 comorbidity (mean = 1.4), and this means that the average patient has a "0" value for all other comorbidities in the model and a final risk adjustment model adjusting for 22 factors.

Because risk adjustment is imperative when measuring functional outcomes, the other measures such as the Competing Measure, NQF #2286 and the Related Measures #2613, #2769, #2775, #2776, #2777 and the FOTO measures (NQF# 0426, 0427, and 0428) also risk adjust for comorbidities.

4. Scale Construct Validity

To ensure strong content and construct validity, the CMS self-care measures only include items related to the construct of self-care, as traditionally defined in functional assessment instruments. CMS recognizes that other aspects of functioning, such as cognition and communication, are important, however, data for these aspects of functioning are typically not aggregated with self-care data to measure improvement in self-care functioning.

Existing literature supports the idea that cognition is a separate construct from motor function (i.e., mobility and self-care) when data from a diverse patient population are analyzed, and concludes that items related to

mobility, self-care, and cognition should not be merged into a combined score (Avlund et al., 1993; Coster et al., 2004; Glenny & Stolee, 2009; Thomas et al., 1998). When selecting the data elements for this self-care measure, our goal was to measure self-care as precisely as possible, and therefore we did not to include items related to cognition.

Feasibility, Usability and Use Considerations

1. Use of Data

The functional assessment data used to calculate this measure will be used by CMS to determine Prospective Payment rates for Medicare part A patients treated in IRFs beginning in October 1, 2019. This data collected for quality measurement are also used for payment. There no costs associated with fees, licensing or other requirements associated with the measure data elements or risk model. All providers have access to a free Java-based software application to collect and maintain their facility's IRF-PAI information. Facilities are able to enter and subsequently export their data from the application for submission to the appropriate national data repository.

2. Interpretability of Performance Score

The performance measure score is presented publicly on IRF Compare as a mean risk-adjusted change in self-care score that is a continuous number and the typical method that IRFs report data. This makes the score more interpretable and transparent to stakeholders and end users. Feedback from Technical Experts in the development of the measures indicated their support for a summed raw item score with the importance of transparency of calculating the quality measure and the ease of data interpretation.

3. Confidential Reports for Providers

Free reports were made available to IRFs through the Certification and Survey Provider Enhanced Reports (CASPER) system starting in 2017. These reports contain feedback on providers' measure performance for internal quality improvement efforts and on national measure scores for quality benchmarking. More details about these reports and what measure data they contain is available in Section 4a2.1.2. under Usability and Use.

4. Public Availability of Measure Data

All measures reported in the IRF QRP serve two purposes: to reflect IRF provider performance by publicly disseminating data about quality of care, which help consumers' and family members' decision making, and to support providers in improving the quality of care they provide to patients. Public reporting on IRF Compare for the functional outcome measures will begin in fall 2020 (on discharges from January 1, 2019 through December 31, 2019).

5. Support for Interpretation and Calculation of Performance Scores

To assist providers to collect accurate data for this measure, CMS has offered multiple in-person and on-line training opportunities since May 2015. In addition, several help desks are available to answer provider questions regarding data collection, and feedback reports, and "Q & A" documents are posted on the CMS website.

To assist providers with calculating their facility's performance score internally, the publicly available IRF QRP Measure Calculations and Reporting User's Manual presents measure specifications and calculations for each measure included in the IRF QRP, including this measure.

To assist consumers, such as family members and patients, with viewing and interpreting the measures posted on the public IRF Compare website, an IRF Public Reporting help desk is available. Individuals can submit questions or comments to CMS at any time and in this way, CMS provides real-time support to patients, families and caregivers seeking additional information or clarification on measures.

References:

Ash, A.S., Shwartz M., Pekoz, E.A., & Hanchate, A.D. Chapter 12: Comparing outcomes across providers. In: Iezonni L, ed. Risk adjustment for measuring health care outcomes, 4th ed. Chicago, IL: Health administration Press; 2012:335-378.

Avlund K, Kreiner S, Schultz-Larsen K. Construct validation and the Rasch model: functional ability of health elderly people. Scand J Soc Med. 1993;21:233–246.

Coster, W. J., Haley, S. M., Ludlow, L. H., Andres, P. L., & Ni, P. S. (2004). Development of an applied cognition scale to measure rehabilitation outcomes. Archives of physical medicine and rehabilitation, 85(12), 2030-2035.

Glenny, C., & Stolee, P. (2009). Comparing the functional independence measure and the interRAI/MDS for use in the functional assessment of older adults: a review of the literature. BMC geriatrics, 9(1), 52.

Mukamel, D. B., Dick, A., & Spector, W. D. (2000). Specification issues in measurement of quality of medical care using risk adjusted outcomes. Journal of Economic and Social Measurement, 26(3, 4), 267-281.

Thomas VS, Rockwood K, McDowell I. Multidimensionality in instrumental and basic activities of daily living. J Clin Epidemiol. 1998;51:315–321.

## Appendix

**A.1 Supplemental materials may be provided in an appendix.** All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

Attachment **Attachment:** IRF_Detailed_Function_QM_Specifications_2633_01-07-2019.docx

## Contact Information

**Co.1 Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services

**Co.2 Point of Contact:** Helen, Dollar-Maples, Helen.Dollar-Maples@cms.hhs.gov, 410-786-7214-

**Co.3 Measure Developer if different from Measure Steward:** RTI International

**Co.4 Point of Contact:** Anne, Deutsch, adeutsch@rti.org

## Additional Information

**Ad.1 Workgroup/Expert Panel involved in measure development**

**Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.**

This quality measure was developed with significant and ongoing input by several Technical Expert Panels (TEPs). Expert panel members provided input on functional status quality metrics, including the performance score, the target population, risk adjustment and exclusion criteria. Some expert panel meetings focused on measuring functional status across post-acute care settings, and other meetings focused on functional assessment and functional outcomes for Inpatient Rehabilitation Facility (IRF) patients.

Most recently, RTI International, on behalf of the Centers for Medicare & Medicaid Services (CMS), convened a Technical Expert Panel (TEP) to seek expert input on the Development and Maintenance of Performance Measures for the Inpatient Rehabilitation Facilities Quality Reporting Program (IRF QRP). An all-day, in-person

TEP meeting was held on March 27, 2017 in Baltimore, MD. The objectives of the TEP meeting were to obtain input on IRF QRP performance measures adopted into the program and obtain guidance and recommendations for future measures. The following experts participated in this TEP:

Mary Ellen DeBardeleben, MBA, MPH, CJCP, Director of Quality at HealthSouth

Karen Green, PT, DPT, Director of Rehabilitation at Cleveland Clinic

Brigid Greenberg, PT, MHS, Business Development Advisor, Manager of Post Discharge Services and Appeals at Uniform Data System for Medical Rehabilitation

Kurtis Hoppe, MD, IRF Medical Director at Mayo Clinic

Cristina Huerta, CRRN, MBA-HCM, Vice President-Rehab Operations, HCA, Inc., Association of Rehabilitation Nurses

Steven Lichtman, EdD, MAACVPR, Patient representative, Director, Cardiopulmonary Outpatient Services, Rehabilitation Research; Research Scientist at Helen Hayes Hospital

Stephanie Nadolny, TRS, MHA, Vice President of Hospital Operations at Spaulding Rehabilitation Hospital Cape Cod

Pam Roberts, PhD, MSHA, OTR/l, SCFES, FAOTA, CPHQ, FNAP, FACRM, Director and Professor Physical Medicine and Rehabilitation and Academic and Physician Informatics at Cedars-Sinai Health System

Mary Van de Kamp, MS/CCC-SLP, Senior Vice President of Quality at Kindred Healthcare

Alan Zaph, PT, Coordinator at Carolinas Rehabilitation – Patient Safety Organization

Previous TEP meetings:

The first expert panel meeting, held as part of a project titled Analysis of Crosscutting Medicare Quality Metrics Using the Uniform Assessment Tool Developed and Tested as Part of the CMS Post-Acute Care Payment Reform Demonstration, was funded by the Assistant Secretary for Planning and Evaluation. The expert panel meeting was held on August 15, 2012, in Washington, DC, with the following expert panel members:

James Farrell, CNO, Healthsouth

David Gifford, MD, MPH, Senior Vice President for Quality & Regulatory Affairs at American Health Care Association

Eileen Bach, PT, M.Ed., DPT, Compliance Specialist, Director Quality and Patient Safety at Visiting Nurse Service of New York

Linda Resnik,PhD, PT, Associate Professor of Health Services, Policy and Practice at Brown University

Trudy Mallinson, PhD, OT, Assistant Professor at University of Southern California, Department of Occupational Science and Occupational Therapy

Margaret Stineman, MD, Professor of Physical Medicine and Rehabilitation, Vice Chair & Director, Research, Department of Physical Medicine & Rehabilitation at University of Pennsylvania

Margaret Rogers, PhD, Chief Staff Officer for Science & Research at American Speech-Language-Hearing Association

Pam Roberts, PhD, OTR/L, CPHQ, FAOTA Manager at Cedars-Sinai Medical Center

Bruce Gans, MDExecutive Vice President and Chief Medical Officer at Kessler Institute

William Pesce, DO, Chief of Physical Medicine & Rehabilitation at Hospital for Special Care

Roger Herr, PT, MPA, COS-C, Vice President Quality Management at Independence Care System

A second expert panel meeting was held on February 8, 2013, as part of a project entitled Symptom Management Measure Development. The following IRF experts were included on this panel:

Alfred Chiplin, JD, Senior Policy Attorney at Center for Medicare Advocacy

Dexanne Clohan, MD, Senior Vice President and Chief Medical Officer at HealthSouth

Cathy Ellis, PT, Clinical Director at National Rehabilitation Hospital, AVP Clinical Services, Spinal Cord Program

Bruce Gans, MD, Executive Vice President and Chief Medical Officer at Kessler Institute

Terrence O'Malley, MD,Medical Director, Non-Acute Care Services

Pamela Roberts, PhD, Manager at Cedars-Sinai Medical Center

Elliot Roth, MD Medical Director, Brain Injury Medicine and Rehabilitation Program at Rehabilitation Institute of Chicago

M. Elizabeth Sandel, MD, Physician

Karen Kloter, Medical Rehab Resource Specialist CARF International

Sharon Sprenger, MPA, RHIA, CPHQ, Senior Advisor, Measurement Outreach, Division of Healthcare Quality Evaluation at The Joint Commission

Suzanne Snyder, MBA, PT, CPUM, Director of Rehabilitation Utilization and Compliance at Carolinas Rehabilitation

Margaret Stineman, MD, Professor of Physical Medicine and Rehabilitation, Vice Chair & Director, Research, Department of Physical Medicine & Rehabilitation, University of Pennsylvania

A third expert panel meeting was held in Baltimore, MD, on September 9, 2013, as part of a project titled Symptom Management Measures. The following experts served on this panel:

Lawrence Miller, MD, Clinical Professor of Medicine at University of California, Los Angeles

Richard Black, MD, Corporate Rehabilitation Consultant at HCR Manor Care

Mary Van de Kamp, MS, CCC-SLP, Senior Vice President of Quality and Care Management at Kindred

Timothy Reistetter, PhD, OTR, Associate Professor at University of Texas Medical Branch

Ellen Strunk, PT, MS, GCS, Consultant at Rehab Resources & Consulting, Inc.

Saad Naaman, MD, MS, Clinician at Physiatry (Physical Medicine & Rehabilitation) Practice

Linda Ladesich, MD, Medical Director Sunflower State Health

Paulette Niewczyk, MPH, PhD, Director of Research at the Uniform Data System for Medical Rehabilitation

Camille Haycock, RN, MS, Vice President, Care Continuum at Catholic Health Initiatives

Elizabeth Newman, OTD, OT/L, Director of Occupational Therapy, Rehabilitation Engineering and Clinical, Informatics at Medstar National Rehabilitation Hospital

Karon Cook, PhD, Research Associate Professor at Northwestern University

Richard Riggs, MD, Chairman and Medical Director for Cedars-Sinai Medical Center

Michelle Camicia, MSN, RN, Director of Operations at Kaiser Foundation Rehabilitation Center

Jill Bolte Taylor, PhD, Author: My Stroke of Insight.

**Measure Developer/Steward Updates and Ongoing Maintenance**

**Ad.2 Year the measure was first released:** 2015

**Ad.3 Month and Year of most recent revision:** 12, 2017

**Ad.4 What is your frequency for review/update of this measure?** annually

**Ad.5 When is the next scheduled review/update for this measure?** 04, 2019

**Ad.6 Copyright statement:** Not applicable

**Ad.7 Disclaimers:** Not applicable

**Ad.8 Additional Information/Comments:** Not applicable